THE ARTICULATION FUNCTIONS AND TEST-RETEST PERFORMANCE OF NORMAL HEARING CHILDREN ON THREE SPEECH DISCRIMINATION TESTS

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY
MARY ELIZABETH SANDERSON

1972

This is to certify that the

thesis entitled

THE ARTICULATION FUNCTIONS AND TEST-RETEST PERFORMANCE OF NORMAL HEARING CHILDREN ON THREE SPEECH DISCRIMINATION TESTS

presented by

Mary Elizabeth Sanderson

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in ~~Audiology~~ & Speech Science

William F. Reinhardt

Date April 27, 1972

ABSTRACT

THE ARTICULATION FUNCTIONS AND TEST-RETEST PERFORMANCE OF NORMAL HEARING CHILDREN ON THREE SPEECH DISCRIMINATION TESTS

By

Mary Elizabeth Sanderson

The purpose of this study was to investigate the performance of a group of normal hearing children on three types of speech discrimination materials: (1) multiple-choice tests requiring no verbal response on the part of the child, (2) conventional tests which have been modified for use with children, and (3) standardized discrimination tests for adults. Tests used to represent these three types of speech materials were the Word Intelligibility by Picture Identification Test (WIPI), the Phonetically Balanced Kindergarten Word Lists (PBK-50), and the Northwestern University Auditory Test No. 6 (N.U. Auditory Test No. 6), respectively. These tests were tape-recorded by a single male talker having a general American dialect.

Subjects were 60 normal hearing children, twelve from each of the age groups: 3½, 5½, 7½, 9½ and 11½. The number of tests and sensation levels administered varied by age group.

The WIPI Test was administered to twelve children with normal hearing from each of these age groups: 3½, 5½, 7½, and 9½ years. The test was administered at 8, 16, 24 and 32 dB sensation levels for all subjects except those aged 3½ years, who were given the test at all but the 8 db sensation level.

It was found that the differences in mean percentage of correct responses were small between adjacent age groups. Mean scores of the 5½ year old subjects were significantly lower than the 9½ year old subjects. (The 3½ year olds were excluded from this analysis.) The mean percentage of correct responses made at the 8 dB sensation level were significantly lower than at the three higher sensation levels. No other paired contrasts of means were significant.

A relatively small number of items on the WIPI test accounted for a large percentage of the errors.

The PBK-50 test was administered to 3½ year old children at a 32 dB sensation level and to older children at the additional sensation levels of 16 and 24 dB. The mean scores of 5½ year old subjects were significantly lower than the mean scores of both 9½ and 11½ year old children. Scores of 3½ year old children fell well below the scores of children aged 5½ years at 32 dB, the only sensation level at which they were tested. All possible paired contrasts between the sensation level means were significant.

The N.U. Auditory Test No. 6 represents the third type of test used, namely, standardized CNC monosyllabic discrimination tests for adults. Children aged 7½, 9½ and 11½ were administered the N.U. Auditory Test No. 6 at 8, 16, 24 and 32 dB sensation levels. Only the sensation level main effect proved significant. The paired contrasts between mean scores obtained at every sensation level were significantly different from each other, with the highest sensation level receiving the highest mean score.

The following conclusions appear warranted. The speech discrimination score obtained for a child at a given sensation level depends in part upon the particular discrimination test used. For 3½ year old children, the WIPI test appears to be the instrument of choice.

For children aged 5½ years, both the WIPI and the PBK-50 tests appear to be appropriate clinical tools. Children aged 7½ years obtain scores similar to children aged 9½ on the WIPI test and to children aged 9½ and 11½ on the PBK-50 and N.U. Auditory Test No. 6.

For all age groups on all three tests the variability of test scores as exhibited by the standard deviation is smallest at 32 dB sensation level. The difference between test-retest scores is small on all three tests. The several lists of each of the three tests are essentially equivalent.

THE ARTICULATION FUNCTIONS AND TEST-RETEST PERFORMANCE OF NORMAL HEARING CHILDREN ON THREE SPEECH DISCRIMINATION TESTS

By

Mary Elizabeth Sanderson

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Audiology and Speech Science

1972

Accepted by the faculty of the Department of Audiology and Speech Sciences, College of Communication Arts, Michigan State University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Thesis Committee:

William F. Rintelmann, Ph. D. Director

Leo V. Deal, Ph. D.

Daniel S. Beasley, Ph. D.

Bernice D. Borgman, Ph. D.

ACKNOWLEDGMENTS

The writer wishes to express appreciation to Dr. William Rintelmann for his guidance as thesis advisor, and also to Dr. Leo V. Deal, Dr. Daniel S. Beasley, and Dr. Bernice D. Borgman for serving as thesis committee members.

Grateful acknowledgment is extended to Dr. Judith P. Frankmann for her assistance in the design of the study, and to Mr. Donald Riggs for his technical assistance.

Special gratitude is also due Mr. Albert Vasils for preparing the figures of this thesis, as well as for his constant encouragement throughout the time of this study.

This work was supported in part by a Traineeship Award from the Rehabilitation Services Administration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Audiological Services for Children

With the rapid development of audiological techniques, including increased sophistication of both tests and equipment, many hearing clinics are providing services to an increasing number of persons every year. A substantial number of young children are included in the population for whom these services are being sought. In the Michigan State University hearing clinic over one-third of the case load was 11 years of age or younger during a ten week period between September and December, 1970.

Several reasons for the large number of children currently being tested can be identified. Increased awareness of the effects on hearing of anoxia (Fisch, 1969), the Rh factor (Flottorp, Morley and Skatvedt, 1957; Markle and Miller, 1963; Flower, Viehweg and Ruzicka, 1966; and Fisch, 1969), maternal rubella (Barr and Lundstrom, 1961; Fisch, 1969; and Miller and Rabinowitz, 1969), and certain other anomalies has led to referrals at an early age and to follow-up testing when one of these conditions is suspected. Cognizance of potential auditory

problems has come from knowledge that hearing impairment is often concomitant with other gross problems and may consequently be camouflaged and go undetected, as in cerebral palsy (Flower, Viehweg and Ruzicka, 1966) and mental retardation (Kodman, 1963). Increased awareness exists on the part of the public that there is an intimate relationship between speech and hearing; so that when normal speech development does not take place, children are more apt to be referred to the hearing clinic to "rule out" hearing loss as a factor. Numerous referrals of children come from physicians who want to determine the amount of hearing loss present during middle ear infections and, later, to evaluate the effectiveness of medication or surgery upon the middle ear. The increasing number of preschool screening programs being established throughout the country has facilitated detection of auditory impairment as well as visual problems at the time of admittance to school. Finally, the expansion of neonatal testing programs results in follow-up tests for children whose early responses to auditory stimulation are questioned (Downs and Hemenway, 1969).

Methods that accurately assess the hearing of children have long been sought. Standard clinical audiometry for both adults and children has routinely included both pure-tone and speech audiometry.

Pure-Tone Audiometry for Children

In testing for pure-tone threshold, the intensity of the auditory signal is gradually increased or decreased until the auditor is able to detect the presence of the stimulus tone approximately fifty percent of the time. This process often proves to be a tedious one for adults, and it has proven to be an extremely difficult task for young children. Many children fail to respond even when a pure-tone signal is known to be substantially above threshold (Froeschels, 1944; Berk and Feldman, 1958; Dixon and Newby, 1959; Barr, 1960; Peterson, 1963; Rintelmann and Harford, 1963; and Clausen, 1966). Explanations offered for this phenomenon include the abstractness of the auditory signal, the short attention span of children, lack of motivation, environmental distractions, rapid fatigue, and sheer boredom at facing a very dull task. In an attempt to allay these problems, a variety of "play audiometry" techniques has been proposed. These techniques include devices such as: the "peepshow" (Dix and Hallpike, 1947; and Dix and Hallpike, 1952), the Pediacoumeter (Guilford and Haug, 1952; and Haug and Guilford, 1960), and conditioning procedures wherein the child places a block in a basket or ring on a peg (Barr, 1955; Lowell, Rushford and Hovarsten, 1956; Geyer and Yankauer, 1957; and Frisina, 1962). Further, some audiologists advocate a selection of whichever technique is felt would be most effective in the

testing of the particular child (O'Neill, Oyer and Hillis, 1961).

Speech Audiometry for Children

Hearing tests employing speech can be classified into two main groups: those that measure speech reception threshold and those that measure speech discrimination. The speech reception threshold (SRT), also called the threshold of intelligibility, is defined as the intensity of stimulus presentation at which an observer can repeat 50 percent of the speech stimuli. Although nonsense syllables, monosyllabic words, connected discourse, and questions have all been used to obtain speech reception threshold, it is current clinical practice to use spondee words (two syllable words with equal stress on both syllables). Several modifications of adult spondee word lists and other new techniques for obtaining speech reception thresholds from children have been suggested (Keaster, 1947; Meyerson, 1947; Monsees, 1953; Sortini and Flake, 1953; Siegenthaler, Pearson and Lezak, 1954; Sims, 1961; and Utley, 1951).

The second type of speech test is the discrimination test. In contrast to the speech reception threshold, testing of discrimination is accomplished at supra-threshold levels. Speech discrimination tests typically contain monosyllabic words with phonetic make-up representing the

frequency of occurrence of the sounds in the English language. The words are presented to the subject at a controlled intensity level, and the percentage correctly reproduced defines the subject's discrimination score.

Significance of the Study

Standardization of tests of speech discrimination in adults has received considerable attention. However, it has generally been agreed that discrimination tests appropriate for adults may not be valid for young children (Nielson, 1960; and Carhart, 1965). Although many tests of discrimination have been proposed for children, standardization of these tests on populations of normal hearing children has received relatively little attention. This oversight is particularly unfortunate because the literature has demonstrated that speech tests provide a more valid estimate of the hearing of many children than do pure-tone tests (Carhart, 1952; Chaklin, Ventry, Barrett and Skalbeck, 1959; Dixon and Newby, 1959; Brockman and Hoversten, 1960; Rintelmann and Harford, 1963; and Clausen, 1966).

When attempting to assess the speech discrimination ability of a young child in the hearing clinic, three types of materials are currently available: (1) standardized discrimination tests for adults, (2) discrimination tests which have been modified for use with children, and (3) multiple-choice discrimination tests requiring no verbal response on the part of the child.

The first two types of tests listed above are "open message" while the third type is a "closed message" type of test. The method or type of test chosen for a particular child is likely to depend upon such subjective criteria as an estimate of the maturity of the child, the extent to which it is felt the child can or will cooperate in the testing situation, and the level of speech and language development of the child. No matter which method of discrimination testing is chosen, there is often an absence of normative data for children, whereas this information has been carefully reported for adult tests of discrimination. Normative data are equally important in interpreting the test results of children.

Purpose of the Study and the Experimental Questions

The purpose of this study was to investigate the performance of a group of normal hearing children on three speech discrimination tests: The Word Intelligibility by Picture Identification Test (WIPI), the Phonetically Balanced Kindergarten Lists (PBK)-50, and the Northwestern University (N.U.) Auditory Test No. 6. The particular tests and sensation levels used with each age group are described in the procedure chapter. The following questions were examined in reference to 3½, 5½, 7½, 9½ and 11½ year old subjects:

1. What is the articulation function, that is, the relationship between percent correct response and sensation level of the stimulus words for each of the three tests?
2. How do the articulation functions of these tests compare with one another?
3. Are discrimination scores obtained from children significantly different as a function of age, test used, or sensation level?
4. How is the variability of test scores related to age and the sensation level of the stimulus?
5. What is the test-retest reliability of each of the tests?

CHAPTER II

REVIEW OF THE LITERATURE

Discrimination Tests for Adults

The foundation of speech audiometry is attributed to the work of Fletcher and his associates at the Bell Telephone Laboratories (Fletcher, 1929). This group was concerned with developing intelligibility lists as a means of determining the efficiency ratings of sound transmission systems. Research begun here was extended at the Psycho-Acoustic Laboratory (PAL) at Harvard University; and it was at PAL that the first intelligibility speech material, also called "articulation tests," were developed for clinical use.

Hirsh (1952) describes an "articulation test" as follows:

> An articulation test is simply a method of measuring the intelligibility of a sample of speech as a function of some variable dimension of the stimulus. The typical procedure involves lists of test items that the listener is to repeat. If he can repeat 45 out of 50 items in a given list, his articulation score is 90 per cent. We can measure the articulation score as a function of many different dimensions of the stimulus. The articulation score obtained for a given set of physical parameters will be different for different test materials (1952, p. 132).

It was decided that monosyllabic words are the most appropriate for determining speech discrimination scores because there is a large pool of English one syllable words from which test items can be drawn. In addition, the words are minimally redundant and hence not too easy, while at the same time they are not as difficult to perceive and repeat as nonsense items.

Egan (1944), at PAL, constructed 20 lists of 50 monosyllabic words with the intention of including phonetic elements in the same frequencies of occurrence they have in the English language. These were called the Phonetically Balanced Lists, or PB-50 lists, and were recorded by Rush Hughes.

At the Central Institute for the Deaf (CID), the PAL lists were revised by Hirsh and associates (1952). Hirsh chose only 200 of the original 1000 words that had been selected by PAL in an attempt to use only those that were most familiar and to achieve better phonetic balance. Hirsh himself was the speaker on the recorded version of this test, called the W-22.

Fairbanks (1958) proposed a new types of test for discrimination testing, namely, the Rhyme test. Drawing a vocabulary from the Thorndike Lorge list, he evolved 50 sets of five rhyming stems differing only in the initial consonant. The purpose of the test was to determine the subject's ability to discriminate initial consonants. Only

18 of 24 English consonants could be used because some consonants do not appear in the initial position in English words and some initial consonant sounds cannot be spelled with a single letter.

Lehiste and Peterson (1959) introduced a new concept in discrimination testing, restricting their test vocabulary to one syllable words both beginning and ending with a consonant. These they called consonant-nucleus-consonant (CNC) lists. In 1962 the CNC lists were revised so that words occurring with a frequency of four per million or less were eliminated (Peterson and Lehiste, 1962).

Tillman, Carhart and Wilber (1963) enumerated the advantages and limitations of the CNC lists for use in determining discrimination scores clinically. The advantages included the fact that each list was made up of only CNC words and each list achieved phonemic balance through comparison with the monosyllables from which the words themselves were drawn instead of from English as a whole. Caution had been taken to be certain the words were as familiar as possible. Limitations of the lists concerned the lack of information on reliability and interchangability of the lists and the fact that Lehiste and Peterson had found it necessary to deviate somewhat from their original plan for phonemic balance.

Using the parent lists of Lehiste and Peterson and conforming to the phonemic balance they suggested, Tillman, Carhart and Wilber developed two lists which

they called N.U. Auditory Test #4, Lists 1 and 2. These were standardized on three groups of adult subjects: 16 with normal hearing, 16 with conductive hearing losses, and 16 with hearing losses of a sensorineural nature.

Even with the six randomizations of the two lists of N.U. Auditory Test #4, it was determined that the repertoire of test items was too meager to permit extensive testing with the same subject. Tillman and Carhart (1966), therefore, revised and expanded this test by constructing four phonemically equivalent CNC lists, called N.U. Auditory Test #6.

These new lists were administered to a group of normal hearing and a group of sensorineural hearing impaired adult subjects for purposes of standardization. It was found that the articulation function of the N.U. Auditory Test #6 was highly similar to that of N.U. Auditory Test #4. Interlist equivalence and test-retest reliability were found to be good. The articulation function for normal hearing subjects on the N.U. Auditory Test #6 was 5.6 percent per dB, falling between the 4 percent per dB of the PB-50 test and the 8 to 10 percent per dB of the W-22 and the Fairbanks Rhyme Test.

Rintelmann and Jetty (1968) used the procedures described by Tillman and Carhart to confirm these findings and to determine whether comparable results would be obtained using a different speaker. Ten normal hearing young adults served as subjects. Test results obtained by

Rintelmann and Jetty were comparable to those found by Tillman and Carhart, with the four lists yielding equivalent articulation functions and little change evidenced between test and retest.

Rintelmann and Burchfield (1970) utilized the four lists of N.U. #6 to test normal hearing subjects and obtained results in good agreement with the findings of Tillman and Carhart.

Several discrimination tests for adults, then, have been devised and standardized. These include the PB-50 test, which has an articulation function of 4 percent per dB; the W-22 test, which has an articulation function of 8 to 10 percent per dB; the Fairbanks Rhyme test with an articulation function similar to that of the W-22 (Kopra, Blosser and Waldron, 1968); and N.U. Auditory Tests #4 and #6, with articulation functions of 6 and 5.6 percent per dB respectively.

Use of Adult Discrimination Tests with Children

The fact that word familiarity is important in the testing of speech discrimination has been demonstrated by a number of investigations of adult subjects. Representative of these are Oyer and Doudna (1959), who found that hearing impaired subjects could discriminate familiar words with greater ease than less familiar items; Owens (1959), who demonstrated that various phonetically balanced word lists of the W-22 test contain a wide spread of

familiarity of test words and that there is a relationship between familiarity and intelligibility of the words; and Rosenzweig and Portman (1957), who found that frequency of word usage is related to intelligibility of words in noise. Results of these and other studies prompted investigation into the appropriateness of using adult tests of discrimination with children.

As early as 1949, Haskins submitted that lists of words familiar to adults (such as the Thorndike and Lorge lists) were not satisfactory in determining words appropriate for the testing of children. Several studies investigated the use of the W-22 discrimination test with children.

Hutton and Weaver (1959) chose the 15 most familiar and 15 least familiar words of the W-22 lists and tested 53 children from the public schools ranging in grade from kindergarten to twelfth grade. It was found that only three of the most familiar words were missed more than once, whereas 14 of the 15 least familiar words were missed more than once. Most of the errors occurred among children in the lower three grade levels. All of the 19 errors on the most familiar words and 75 percent of the errors on the least familiar words occurred at kindergarten and first grade levels. The authors further found that the least familiar words did not distinguish between listeners in the same manner as did the most familiar words. The explanation for this effect suggested by the authors was that

the least familiar words test the vocabulary as well as the sound discrimination ability of the client. The authors stated that their test results ". . . cause serious concern about the use of the PB W-22 lists and recordings in hearing testing at the lower age levels" (p. 1448). They further suggested that "these data may also be a partial explanation for the variability encountered in some clinical situations" (p. 1449).

Brooks and Goetzinger (1966) suggested interpreting the findings of Hutton and Weaver with caution because they felt that a marked dichotomy between familiar and unfamiliar items is not present in actual clinical application. Brooks and Goetzinger administered recordings of the Rush Hughes PB-50 List 7 and the W-22 List 2 to groups of normal hearing children in grades 2, 4 and 6. Both word lists were analyzed with reference to three vocabulary characteristics: (1) multiordinality, (2) abstractness, and (3) familiarity. Significant differences in multiordinality between the two lists were found, as well as differences in word familiarity. However, these differences were not found to be related to intelligibility, except for "a low correlation between discrimination difficulty and word familiarity when based only on very young children (Grade 2)" (p. 367). Significant differences between the W-22 List 2 and List 7 of the Harvard PB words were found at each grade level, but significant differences between grades were not

evidenced for either test. Brooks and Goetzinger stated that although analysis of discrimination score differences by grades does suggest an age trend, this should "in no way hamper the usefulness of the tests with children . . . if normative data for all ages could be developed" (p. 367). By Grade 6, the authors point out, the means from their study (94.75) closely approximate those reported by Ross et al. (1965) and Giolas and Epstein (1963), who obtained mean scores of 96.0 and 95.7 percent respectively utilizing the same W-22 lists on adult subjects.

Using as subjects a group of normal hearing college adults and normal hearing children in grades 1 through 10 (excluding second grade), McNamee (1960) found that the ability of children to respond correctly to the CID Auditory Test W-22 decreased with decreasing age.

Nielson (1960) analyzed the results obtained from preschool children responding to the CID Auditory Test W-22, using as subjects normal hearing children ranging from 3.0 to 5.8 years of age. She found a "systematic difference" in the performance of 3, 4 and 5 year olds on the CID Auditory Test W-22, with the range of deviation becoming smaller as age increased. Correlation values between age and score within each age group indicated a low, positive correlation between age and test score, with a tendency for this correlation to lessen with increased age.

In summary, several studies utilizing adult tests of speech discrimination with normal hearing children have been reviewed. These studies include that of Hutton and Weaver, who found that the 15 most familiar and 15 least familiar words of the W-22 did not distinguish between children in the same direction. They further reported that all of the errors on the most familiar words and 75 percent of the errors on the least familiar words occurred in Kindergarten and first grade children. Brooks and Goetzinger, using both the Rush Hughes PB-50 recording and the W-22, and testing children in grades 2, 4 and 6, found a correlation between word familiarity and intelligibility with children in Grade 2 only and also found that scores of children in Grade 6 approximated that of adults. McNamee, comparing adult scores with scores of children in grades 1 through 10, not including Grade 2, found a significant difference between scores of adults and scores of children in grades 1, 3, 4, 5, 6 and 8, using the W-22. The W-22 was also utilized by Nielson, who found a "systematic difference" between the scores of children aged 3, 4 and 5, with the range of deviation in scores becoming smaller as age increased.

Modified Discrimination Tests for Children

Haskins (1949) constructed a test to meet a need for "discrimination tests which are appropriate for use

with hard-of-hearing children and with adults whose language background is limited." Her goal was to develop an appropriate test for children by simplifying the vacabulary of the PB-50 lists while maintaining the basic design of these same lists. Haskins first catalogued the original pool of 1000 words comprising the PB-50 test in terms of phonetic composition. She then evaluated the words in terms of their vocabulary levels, ultimately choosing only those words that also appeared on the International Kindergarten Vocabulary List. This particular list of words was felt to be most appropriate because it is composed of words observed to be present in the speech of children before they enter first grade.

A total of 425 monosyllabic words was selected; and items were chosen from this group to make up four sets of fifty items each, using no word more than once and following Egan's pattern of phonetic balance as closely as possible. Haskins called the resulting word sets the PBK or PBK-50 lists.

In order to determine their relative difficulty, Haskins plotted articulation functions for 22 normal hearing adult subjects on each of the PBK-50 lists and List 13 of the Harvard PB-50 test. Analysis indicated that Lists 1, 3 and 4 of the PBK-50 test were equivalent to List 13 of the original PB-50 test, but PBK-50 List Number 2 was easier than the others.

In an attempt to estimate the reliability of the PBK-50 test materials and procedures, each list was presented to the same adult subjects and a retest was administered at one sensation level. It was found that there was a "consistent trend for better scores to appear on the retest." Haskins attributed this to a learning factor. She wrote:

> On the basis of other findings, one would not expect the reliability of the PBK-50 Test to deviate markedly from that for the original PB-50 Test. However, this matter must be explored anew by doing test-retest measurement throughout a range of presentation levels covering the full articulation function for each list (1949, p. 70).

It should be pointed out, however, that the PBK-50 test was not standardized with either normal-hearing or hard-of-hearing children.

Hudgins (1949) reported a test developed by selecting four lists of 50 words from the original PB lists developed at PAL. The words were chosen for their familiarity and their phonetic content. Each list was composed of speech sounds in the proportion in which they occur in English prose, and the lists were equated for relative difficulty. Although Hudgins originally intended the lists for use in evaluating expressive speech of hearing-impaired children, the lists have subsequently been used for auditory speech discrimination testing of children, and are called the PBF lists (Phonetically

Balanced Familiar Words). Very few published studies have been reported utilizing this test.

Multiple Choice Discrimination Tests

Sortini and Flake (1953) described a technique used by them for assessing the ability of preschool children to hear and understand speech. The child is shown several toys, and speech reception threshold is established by asking the child to do such things as "Show me the airplane," or "Give mother the baby." The child's discrimination score is obtained by shortened forms of toy names (e.g., "plane" for "airplane"). Sortini and Flake implied reliability when they concluded:

> Thus it would seem that the test described which (1) requires short administration time, (2) insures a high degree of subject cooperation, and (3) holds the child's attention for the whole testing period, is a reliable evaluation of the preschool child's ability to hear and understand speech from an active communication standpoint. This eliminates the need at this age level for the more time-consuming and less reliable pure-tone audiometric test (1953, p. 997).

Keaster (1947) was the first to submit a multiple-choice picture test for use in determining the hearing status of young children between the ages of three and six. However this was a speech reception threshold test rather than a discrimination test. She constructed the picture type test in an effort to devise a test that would:

1. Appeal to children.
2. Require no verbal response on the part of the child.
3. Take into account the short memory span of the very young.
4. Be short enough to hold the attention of even the youngest in the age group.
5. Use a standardized vocabulary (1947, p. 159).

Landes (1960) suggested a need for this type of picture hearing test for purposes of assessing discrimination in children. He wrote:

> If the child is immature or has an articulation defect, it is impossible to evaluate an erroneous response to a PB stimulus word in terms of whether he failed to hear the stimulus correctly, or whether he simply cannot repeat it correctly. If the expected responses were nonverbal in character, an erroneous response could be construed as a result of a defect of hearing. . . . (1960, p. 248)

Sims (1961) constructed a picture-identification multiple-choice test of twenty-five phonetically balanced, monosyllabic words, as well as a picture identification test for use with children for the measurement of speech reception threshold. The suitability of the test material for use with young children was evaluated upon a population of children between 3 and 6 years of age. Sims concluded that the test was unsuitable for use with three year old children and that revision of the pictured items was needed before the test would be suitable for use with four and five year old children. To evaluate test validity, a sample population of adults with hearing loss was utilized. The discrimination score of the adults based on Sims' test were compared with the scores obtained through use of

conventional adult tests. Sims wrote that the results indicated some revision of the test structure was needed.

Seidel (1963) constructed a picture test of auditory discrimination for young children, choosing 174 monosyllabic nouns from dictionaries, children's vocabulary lists, and other test materials. Her criteria for selection of words included (1) familiarity to 3, 4 and 5 year old children, (2) the ability of the word to be represented by a simple picture, and (3) the feasibility of placing the word in a triplet of two other words with the same medial vowel sound. An attempt was made to include a wide representation of the phonemic elements of speech. All except four of the 174 monosyllabic nouns chosen were of the CNC variety. Seidel called this the PICSI Test: Picture Identification for Children--A Standardized Index. Nasca (1964) applied the PICSI test of Seidel to a population of children with sensorineural hearing losses and delayed language in an effort to facilitate early diagnosis of hearing problems.

Haspiel (1961) submitted a new concept in discrimination testing, suggesting that phonetic balance may be of "relatively small importance." He suggested that factors other than phonetic balance should be considered: (1) familiarity of vocabulary; (2) phonemic confusion of test words; and (3) the phonetic factors of voicing, influence, and pressure pattern.

In a discussion of the phonetic factors, Haspiel wrote:

> Because these three factors seem to be the essential events necessary to describe a given phoneme spectrographically, it would appear logical that they might also be the three essential factors necessary for an analysis of audible speech. It was therefore decided to vary these three phonetic factors systematically in the construction of a picture identification test of discrimination for audiiological purposes.
> . . . . . .
> The child was asked to discriminate between a series of paired picture-words in which the following categories of acoustic differences are systematically varied: (V) voicing, (I) influence, (P) pressure pattern, (VI) voicing and influence, (VP) voicing and pressure pattern, (IP) influence and pressure pattern, and (VIP) voicing, influence, and pressure pattern (1961, p. 44).

An initial vocabulary was chosen by Haspiel from the Thorndike-Lorge list. Words were then paired on the basis of phonetic differences between initial consonants only, the middle and final phonemes of each word-pair being identical. Care was taken to select words that (1) occurred at least one time in every million words, (2) could be represented through pictures, (3) were monosyllabic cognates, and (4) were CNC combinations utilizing no blends. A set of 48 word-pairs was selected on the basis of familiarity to young hearing impaired children and their discriminating ability when presented to 20 normal-hearing, 20 sensorineural-impaired, and 20 conductively-impaired children between the ages of 9 and 12

years. The 48 word-pairs chosen were called the Picture Identification Test of Discrimination (PITD).

Empirical validity of the test was determined by establishing its relationship to the PBK-50 word list. A moderate, positive relationship was discovered between the PBK-50 test and the PITD. This moderate correlation was interpreted by the author to indicate that the PBK-50 test and the PITD measure "similar but not identical aspects of auditory ability. This result is reasonable in view of the differing rationale for selection of items which constitute the two tests." The author further concluded that "it is possible to predict the performance of normal or hearing impaired subjects from one test to the other with an error of plus or minus about 10%."

Siegenthaler and Haspiel (1966) conducted further research to standardized the speech intelligibility materials proposed by Haspiel as well as speech reception threshold materials. Three forms of the test of 48 word-pairs were devised by selecting the stimulus word for the first list by chance, selecting the alternate word for the second list, and randomly selecting one of the two stimulus words of each word-pair for the third list. The name of the discrimination test was changed from PITD to DIP (Discrimination by Identification of Pictures).

As part of this research, 295 children with normal hearing between the ages of two years ten months and eight

years three months, inclusive, were tested. Children were retested within one week of the initial test. During each test session, each child received one form of the DIP at a 0 dB sensation level, one form at a 5 dB sensation level, and the third form at a 10 dB sensation level.

Test results indicated that the three forms of the DIP were essentially equivalent and that boys and girls in the age range studied had essentially the same DIP scores with no difference in variance of the DIP scores between the sexes. Test performance for both males and females improved with increased age over the range of three years to eight years, with an increase of about nine in 48 items. There was an indication of "leveling off" at the upper age limits.

In reference to test-retest reliabiltiy, the authors wrote: ". . . for the present the estimate of test realibility (the range within which two-thirds of the DIP test-retest differences would be expected to fall) is ± 5 items" (p. 81).

Myatt and Landes (1963), in constructing a picture discrimination test for children, chose eighty pictures representing monosyllabic words judged to be within the vocabulary of preschool children. They assembled the words into twenty groups of four words each. They stated:

> Some groups, e.g., "fish-cup-blue-book," were specifically grouped to represent phonetically dissimilar words, whereas other groups, e.g.,

> "coat-cone-Coke-comb," were chosen to portray words of marked phonetic similarity. A third type of grouping included elements of both similarity and dissimilarity, e.g. "boy-wing-watch-ring (1963, p. 360).

The test was administered to ten children from each of the following four groups: a group with normal intelligence and speech, a speech-defective group, an educable mentally retarded group, and a trainable mentally retarded group. All children had normal hearing.

In comparing the four groups it was found that the educable mentally retarded group had the smallest number of errors. The authors explained this by pointing out that the curricula to which these children were exposed was highly "picture-oriented." The scores of the trainable mentally retarded group were significantly lower than those of the others, but there were no significant differences in the test results of the other three groups. The authors concluded:

> The picture test is presumed to yield essentially the same informatin available through the use of PB lists which are practical with an adult population but not with some children. Further study is indicated to explore the validity of this test as a test of discrimination loss (1963, p. 362).

Lerman, Ross and McLauchlin (1965) employed an artist to paint the 80 pictures of the Myatt and Landes test described above and then administered the test to 30 hearing impaired children ranging in age from four years eight months to ten years eleven months at as close to a 40 dB sensation level as possible. The

authors concluded that the Myatt and Landes test is capable of assessing the discrimination score of the hearing-impaired children in a reliable and comparable manner.

Ross and Lerman (1970) revised and extended the above test in an effort to remedy problems encountered in its administration. The shortcomings they specifically sought to eliminate included the following: (1) some of the pictures were inadequate representations of the words, (2) some of the words were too difficult, and (3) chance scores were too high. Ross and Lerman called the revised edition of the test the Word Intelligibility by Picture Identification Test (WIPI). Stimulus words included words on the previous test which had been found satisfactory, supplemented by monosyllables found frequently in children's books and word-count lists. New words considered questionable were presented to kindergarten and nursery school teachers for judgment as to their appropriateness. A preliminary evaluation of the test vocabulary was accomplished with a group of fifteen hearing-impaired children ranging in age from 6 through 12.

The stimulus words were arranged in 25 sets of six words each, representing an expansion of the original 20 sets of four words each. As in the previous versions, some matrices presented gross discrimination tasks as well as other discrimination tasks of a finer nature. Four lists were compiled by choosing one stimulus word from each matrix for each list. In this manner no stimulus word was used in

more than one list, and two words of each matrix were not used in any of the lists but acted as additional foils. There was no attempt to attain phonetic balance. They stated:

> The equalization of the test lists was accomplished partly on the basis of our experience, partly on the basis of acoustic phonetic considerations (Liberman et al., 1967), and partly on an a priori basis. In the final evaluation phase our judgments underwent empirical verification (1970, p. 49).

A final evaluation of the test involved 61 subjects, none of whom had been used in the preliminary stage of the study. They ranged in age from four years seven months to 13 years nine months, with a mean age of ten years two months. All subjects had hearing losses exceeding 30 dB (re: ISO 1964) at one or more of the speech frequencies, with a two-frequency pure-tone average (PTA) in the test ear ranging from 5 to 90 dB, with an average of 52.2 dB. Fifty-eight subjects evidenced a congenital sensorineural hearing loss, and three had long-standing losses which were conductive or mixed in nature. Twenty-four of the subjects were enrolled in a school for the deaf.

All four test lists were presented via live-voice to each subject at a 40 dB sensation level (re: the two-frequency PTA), or, if this was not feasible, 5 dB below the uncomfortable loudness level (UCL) of the subject. The lists were re-administered to the subjects between one and four weeks later, at the same sensation level. Test-retest reliability coefficients ranged from 0.87 to 0.94, and the errors of measurements from 4.7 to 7.7. The authors wrote:

"These results indicate that all four lists of the test are highly reliable, with comparable reliabilities for all four lists" (p. 50). Differences between test and retest presentations indicated that a learning effect had taken place. Differences in the second presentation were found to be not more than the equivalent of one extra word correct, however, and were not considered clinically significant.

In reference to the equivalence of the lists, the authors concluded:

> A comparison of the means and standard deviations of the four lists indicates that the average level and range of difficulty are comparable. The only significant mean difference was at the 0.05 level and occurred between lists III and IV. However, this difference of 2.8% is less than a one-word variation and, as in the learning effect, cannot be considered to be clinically significant (1970, p. 51).

Equivalence of the lists was considered to be further supported in the Pearson product-moment correlation coefficients of the four lists, which ranged from 0.84 to 0.95. Five of the six correlations were 0.92 or higher. Pearson product-moment correlation coefficients between the list scores and the PTA's showed correlations ranging from 0.60 to 0.65 and were negative, indicating that subjects with less hearing loss had better scores on the discrimination task.

The authors concluded: ". . . it appears that the test is suitable for children with moderate hearing losses from ages five or six and for children with severe hearing losses from ages seven or eight" (p. 52). They further wrote:

> Based on theoretical expectations and the limited data we do have, discrimination scores obtained with the picture identification test should exceed conventional test scores by approximately 25%. A direct evaluation comparing scores obtained on the present test with scores obtained on different types of tests would be a useful project for a future investigator to undertake.
>
> The design of the study did not permit the determination of articulation-gain functions. These data would permit a more complete description of the properties of the test, and it would be useful to obtain this information in future research (1970, p. 52).

Summary

Several speech discrimination tests for children have been reviewed. Some have been designed with modified vocabularies for use with children, have attempted to retain phonetic balance similar to that of adult tests of speech discrimination, and require verbal responses from the child (Haskin, 1949; and Hudgins, 1949). Other discrimination tests are of the multiple-choice variety (Sims, 1961; Haspiel, 1961; Myatt and Landes, 1963; Seidel, 1963; and Ross and Lerman, 1970).

Of all of these tests, only the Discrimination by Identification of Pictures (DIP) has been standardized with normal hearing children, including determination of articulation functions (Siegenthaler and Haspiel, 1966). Articulation functions have not been established for any other discrimination test for either normal or hearing-impaired children.

CHAPTER III

EXPERIMENTAL PROCEDURES

Subjects

The subjects of this study were 60 normal hearing children, divided equally between the ages of 3½, 5½, 7½, 9½ and 11½ years. Thus, there were 12 children in each age group. Only children within three months of the ages stated above were tested. These ages were chosen to represent samples of the age levels of clients frequently tested in hearing clinics and for which little normative data have been established.

Mean ages of the five groups in months were 42.1, 65.9, 88.8, 113.2 and 137.2. In the total sample there were 26 boys and 34 girls. Specific information as to the age and sex of each subject can be found in the Appendix, along with scores on each of the tests administered. Pure-tone screening audiometry was employed for determining that the hearing of the subjects was within normal limits. The details of this screening procedure are described later in this chapter.

Prior to administering pure-tone and speech audiometry, the experimenter conversed with all children to be tested in order to establish that the speech of the child fell within normal limits for his age group.

One 5½ year old subject had to be omitted when he refused to repeat any of the spondee words. An 11½ year old subject developed a cold between test sessions and had to be rejected because of a pronounced shift in her speech reception threshold. In addition, one rather hyperactive 9½ year old youth was returned home after he twice attempted to dismantle a microphone prior to testing.

The equipment and procedure utilized in recording and balancing the stimulus materials will be described under test materials.

For speech testing, all of which was done in sound field, a commercially available speech audiometer (Grason-Stadler, Model 162) was used to amplify and attenuate the electrical output of the tape recorder (Ampex, Model 500) used to present the tape recorded tests to be described under test materials. The output of the speech audiometer drove a single Grason-Stadler, Model 162-4 loudspeaker.

The speech audiometer was calibrated so that audiometric zero was defined as being 13 dB above 0.0002 micro

bar. "Speech noise" was used for calibration, in the manner described by Tillman, Johnson, and Olsen (1966), who reported that the spectral configuration of "speech noise" is closer to the spectrum of speech produced by male speakers than is a 1000 Hertz tone.

Calibration of the loudspeaker was accomplished by placing the condenser microphone (Bruel and Kjaer, Type 4131) four feet from the face of the loudspeaker at a height of 42 inches. The condenser microphone was positioned so that its diaphragm was perpendicular to the floor and ceiling of the test chamber at a zero degree angle of incidence from the loudspeaker. The intensity of the speech spectrum noise generated by the speech audiometer at a 60 dB attenuator setting was then recorded. All measurements were made without the presence of an observer in the field; however, the location of the condenser microphone was approximately where the center of the subject's head would be when a subject was in the test chamber. A pistonphone (Bruel and Kjaer, Type 4220) was used to set the meter needle of the audio frequency spectrometer (Bruel and Kjaer, Model 2112) from which the intensity of the sound-field was read directly in decibels re: 0.0002 microbar. Measurements of the overall SPL of the speech noise were made on all days that subjects were tested, and adjustments made so that the readings were within plus or minus one decibel of 13 dB re: 0.0002

microbar. Attenuator linearity was also checked throughout the experiment.

For pure-tone screening tests, a commercially available pure-tone audiometer (Beltone, Model 15C) was used to drive TDH-39-10Z transducers housed in MX 41/AR biscuit-type cushions. The pure-tone air conduction system was checked periodically, using a sound level meter (Bruel and Kjaer, Type 2203) with its associated octave band filter (Bruel and Kjaer, Type 1613), an artificial ear (Bruel and Kjaer, Type 4152) and a condensor microphone (Bruel and Kjaer, Type 4132).

Test Environment

The test room and all audiometric equipment were located in the Michigan State University Speech and Hearing Clinic. The earphones used for pure-tone screening and the loudspeaker used in speech audiometry were situated in a prefabricated double-walled test room (IAC Series 1200) of this building. The pure-tone audiometer, speech audiometer, and tape recorder were located in an adjacent single-walled control room (IAC Series 400), communicating with the test room by means of a window and a two-way electronic communications system. The subjects were seated in the test room at a zero degree angle of incidence from the loudspeaker for all speech audiometry.

Test Materials

The following stimulus materials were utilized in this study:

The Utley Children's Spondees, Randomizations A, B and C
The PBK-50, Lists 1, 2, and 3, Randomizations A and B
N.U. Auditory Test No. 6, Lists 1, 2, 3 and 4, Randomizations A and B
The Word Intelligibility by Identification of Pictures (WIPI) Test, Lists 1, 2, 3 and 4, Randomizations A and B

Two forms of all of the above tests were utilized except for the Utley Children's Spondees, of which three forms were used. The method used in devising and assembling the various test forms will be described later.

The Utley Children's Spondees consist of a group of two-syllable simple words intended for obtaining speech reception thresholds with young children.

The PBK-50 test is made up of three phonetically balanced lists of 50 words each, all of which appear on the International Kindergarten Vocabulary List and have been found to be present in the speech of children before they enter first grade.

The N.U. Auditory Test No. 6 represents an expansion of the N.U. Auditory Test No. 4. It is comprised of four equivalent, phonemically-balanced CNC lists of 50 words each.

The WIPI Test is a multiple-choice picture discrimination task made up of four lists of 25 words each. Six pictures appear on each page of the test booklet. Four of

the pictures represent stimulus items, one from each of the four lists, with the remaining two pictures acting as foils. No attempt was made to achieve phonetic balance within these lists.

All stimulus material was recorded by the same male speaker, with a General American Dialect, and experienced in monitored live-voice speech audiometry. During all recording he spoke into an Electrovoice 635A microphone at an intensity monitored to cause a deflection to 0 dB on a VU meter. During the recording of the Utley Children's Spondees, both syllables of the stimulus words were monitored. For the PBK-50 and N.U. Auditory Test No. 6 the final word of the carrier phrase "You will say . . ." was monitored with the test item allowed to fall naturally. The monitored carrier phrase for the WIPI test was "Show me . . ." with the stimulus item again permitted to follow naturally.

The initial recordings were made on an Ampex Model AG 500 tape recorder, with the speaker in a single-walled, sound treated booth (IAC, Series 400). The tapes of these recordings were balanced utilizing a level recorder (Bruel and Kjaer Model 2305) and dubbed onto another tape. The output of all of the spondee words and the last word of each carrier phrase were equated to within ±2 dB of the level of a 1000 Hertz calibration tone spliced to the beginning of each tape. The Ampex Model AG 500 tape recorder, in conjunction with an Ampex AG 602 tape recorder, were utilized for this purpose.

The balanced original copy of the words was considered to be Form A of all lists. Form B of all discrimination materials and Forms B and C of the Utley Children's Spondees were devised in a manner similar to that reported by Tillman and Carhart (1963 and 1966). Form A of the Utley Children's Spondees was copied twice since two additional forms were required, and Form A of the discrimination stimulus materials was copied once. The duplicate copies of the lists were reorganized into different scramblings (forms) by means of cutting the magnetic tapes into numbered segments and splicing them together in an order determined by a table of random numbers. In this way stimulus equivalence between the randomizations was achieved. Five seconds of leader tape was spliced between items to maintain a constant temporal pattern.

Test Procedure

Each of the sixty subjects was given a pure-tone air conduction screening test bilaterally at the beginning of the first test session. The screening test was administered at a 20 dB hearing level re: ISO 1964, at the following frequencies: 250, 500, 1000, 2000, 4000 and 8000 Hertz. The 3½ and 5½ year old subjects were conditioned to put a block on a peg for purposes of screening. The older subjects were simply asked to raise their hand upon hearing the pure tone signal. Any subject who did not respond to all frequencies in each ear was eliminated.

Speech audiometric tests were administered to each subject on either two or three separate days. In each test session a sound-field speech reception threshold was obtained and was used as the basis for administering the speech discrimination tests. Three tests were employed to measure speech discrimination for the older subjects: (1) the WIPI, (2) the PBK-50, and (3) the N.U. Auditory Test No. 6. Subjects aged 3½ and 5½ years were not given the N.U. Auditory Test No. 6.

In order to obtain articulation functions for the three speech discrimination tests, each test was given at several sensation levels for most age groups. In determining the articulation scores for every test, each successive presentation of the test list was given at a higher intensity, with one list of the test given at each sensation level tested. The specific sensation levels used, as well as the tests administered to each age group, are indicated in Table 1.

The decision to limit the number of sensation levels tested for the younger children was based on the necessity of maintaining interest in the task and minimizing fatigue. All discrimination tests were followed by a retest, administered approximately 10-15 minutes following the test session. The order of test administration as well as the order of list presentation within each test was determined by rotation. For the retest, the alternate

TABLE 1.--The sensation levels at which children of each age group were administered the WIPI, PBK-50, and N.U. No. 6 speech discrimination tests.

Age	Test	8	16	24	32
3½	WIPI		x	x	x
	PBK-50				x
5½	WIPI	x	x	x	x
	PBK-50		x	x	x
7½	WIPI	x	x	x	x
	PBK-50		x	x	x
	N.U. #6	x	x	x	x
9½	WIPI	x	x	x	x
	PBK-50		x	x	x
	N.U. #6	x	x	x	x
11½	PBK-50		x	x	x
	N.U. #6	x	x	x	x

randomization of the same list of words as was used previously at that sensation level was utilized.

Only one discrimination test (together with the retest) was given to a particular subject on any one day. All testing for a particular subject was completed within one week of the date the initial test was administered.

The speech reception threshold of each subject was established at the beginning of every test session in the manner described by Tillman, Carhart, and Wilber (1963).

This method includes familiarizing the subject with the spondees in advance of the testing by means of having the subject repeat the stimulus words as they are read to him in a face-to-face situation. The following instructions were given:

> I am going to say some words. I would like to have you say the same words after me . . .
> . . . . . .
> Now you will hear a man saying the words. You will hear the words coming from here (point to the speaker). Some of the words will be very soft. Listen carefully, and say the word you think the man said. I will be sitting in the next room, but I can hear you.

The spondees were then presented to the subject in descending 2 dB steps,beginning at a sensation level of 10 to 20 dB above the anticipated threshold. Four words were presented at each step until the lowest level at which either two out of the first three or two out of four spondees were repeated correctly. This level was designated as the speech reception threshold. The stimulus material used in establishing speech reception threshold in this study was the Utley Children's Spondee List.

Instructions for the picture test of discrimination (WIPI) were as follows:

> Here are some pages that have pictures on them. In a few minutes you will hear a man talking to you. The man is going to tell you what picture to show me on each page. Look at the pictures on the page, then point to the picture he tells you to show me. Let's try it first with this page (practice sheet).
> Show me cat.
> Show me glass.
> Show me rat.
> Now let's go on to the next page. Look at the pictures, and listen, and the man will tell you what to show me.

The tester remained in the testing suite with the child during the administration of the WIPI in order to assist in turning the pages of the test booklet and to insure accurate scoring of test items. During this test the experimenter controlled the timing of the stimulus items by means of a remote control unit in conjunction with the Ampex AG 500 tape recorder.

For the threshold testing, and during the administration of the PBK-50 and N.U. Auditory Test #6, the experimenter remained in the control room and the subjects, except for the 3½ year olds, were alone. A parent or sibling of the 3½ year old subjects was in the test room with the subject during testing. Instructions for the PBK-50 and the N.U. Auditory Test #6 were as follows:

> Now you will hear some different words. Listen and say the word I tell you to say. 'You will say pack.' 'You will say 'eat.' You will say 'map.'
> Now you will hear a man telling you to say some words. Listen, and say the word he tells you to say. If you're not sure, say the word you <u>think</u> he said.

Recall, only one discrimination test, together with the retest, was given to a particular subject on any one day.

CHAPTER IV

RESULTS AND DISCUSSION

This chapter is divided into four sections. The first three sections present the primary test results and discussion of the WIPI, the PBK-50, and the NU Auditory Test No. 4, respectively. The fourth section of the chapter presents a general discussion of the tests as they relate to each of the age groups used in this study.

The Word Intelligiblity by Picture Identification Test (WIPI)

The WIPI test, it will be remembered, is a multiple-choice picture discrimination test made up of four lists of 25 monosyllabic words. One word from each of the four lists is represented pictorially on each page of the test book, together with two additional pictures which are not stimulus items but act as foils. Each matrix, therefore, contains six pictures, some presenting gross discrimination tasks and some discrimination tasks of a finer nature.

Twelve children with normal hearing from each of the age groups 3½, 5½, 7½ and 9½ were administered the WIPI test in this study. The 3½ year old subjects were tested

at sensation levels of 16, 24 and 32 dB. The older subjects were given an additional presentation at an 8 dB sensation level. A different word list was used at each presentation level for the test session. A retest followed within 15 minutes of the completion of the test session, at which time the lists were again presented to the subject at the same sensation level at which they had been administered during the test session. The 3½ year old subjects were retested only at a 24 dB sensation level.

Table 2 displays the data obtained for the subjects during the test session. Table 3 summarizes the same information obtained in the retest session. In these two tables the means and standard deviations are indicated for each age group at every sensation level utilized in the test and retest sessions. The standard error of measurement for each sensation level is included in Table 2.

The mean values reported in these two tables are displayed in graphic form in Figure 1.

WIPI Articulation Function

For purposes of plotting the articulation function, all four WIPI lists were combined. Examination of Figure 1 reveals that the articulation functions yielded by the children aged 5½, 7½ and 9½ are highly similar to one another in configuration. It can also be observed that the slope of the articulation function changed little between test and retest for any age group.

TABLE 2.--Mean (M), standard deviation (SD) and standard error of measurement (Se) in percent correct of 12 normal hearing children from each of the age groups 3½, 5½, 7½ and 9½ on the WIPI during the test session.

Age	Sensation Level in dB	M	SD	Se
3½	8	--	--	--
	16	78.68	10.14	--
	24	88.32	10.85	--
	32	91.68	6.26	--
5½	8	83.00	8.20	9.74
	16	94.00	7.91	4.36
	24	95.68	3.17	3.10
	32	97.32	2.16	2.11
7½	8	89.32	8.56	6.94
	16	95.32	4.56	2.58
	24	97.32	2.61	2.51
	32	98.68	2.61	1.45
9½	8	89.32	7.88	5.46
	16	96.32	3.17	2.17
	24	98.68	1.97	1.82
	32	99.00	1.80	0.96

TABLE 3.--Mean (M) and standard deviation (SD) in percent correct of 12 normal hearing children from each of the age groups 3½, 5½, 7½ and 9½ on the WIPI during the retest session.

Age	Sensation Level in dB	M	SD
3½	8	--	--
	16	--	--
	24	87.67	12.58
	32	--	--
5½	8	87.67	11.37
	16	95.67	5.25
	24	95.33	7.20
	32	98.33	3.17
7½	8	90.33	8.44
	16	97.33	1.97
	24	98.67	2.61
	32	99.00	2.49
9½	8	90.33	5.25
	16	97.33	2.61
	24	99.00	1.81
	32	99.33	1.56

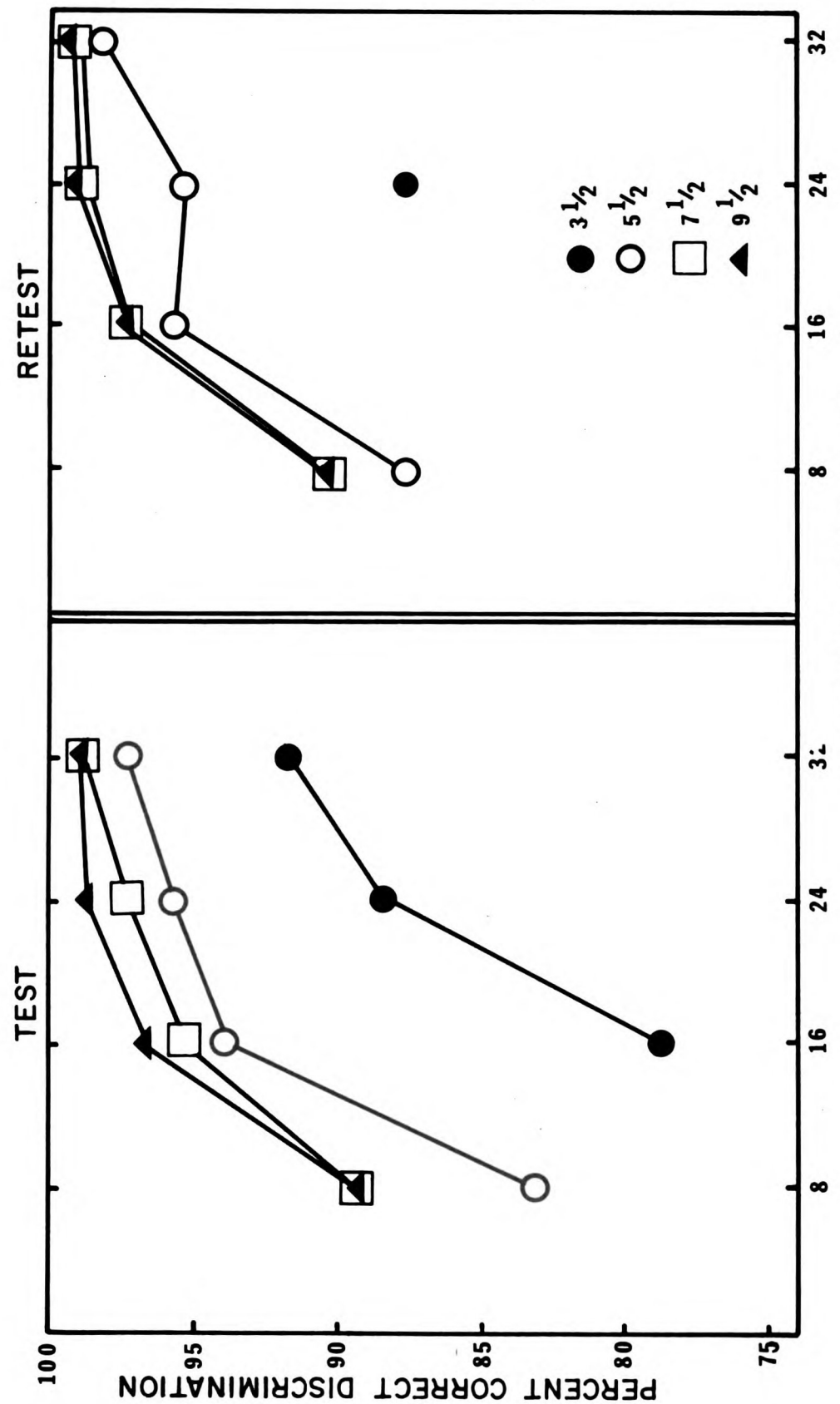

Figure 1.--The mean discrimination scores of 12 normal hearing children from each of the age groups 3½, 5½, 7½, and 9½ on the WIPI during test and retest sessions.

Tables 2 and 3 reveal that, although the differences in the mean number of items correct between adjacent age groups are small, the mean scores are positively correlated with the age of the subjects. The mean scores of the 9½ year old subjects equal or exceed those of the 7½ year old subjects at every sensation level, for both test and retest sessions (although on retest the differences are so slight as to be clinically negligable). The scores of the 7½ year old children exceed those of the subjects aged 5½ at every sensation level. Scores of the 3½ year old subjects fall substantially below their 5½ year old counterparts.

A two way analysis of variance for repeated measure design, summarized in Table 4, confirms that the effects associated with both age and sensation level are, indeed, statistically significant. The 3½ year old subjects were not included in this analysis because they were not tested at the 8 dB sensation level.

The Tukey post hoc method of multiple comparisons was used to test the significance of individual treatment means for age and sensation level. As shown in Table 5, the mean scores of the 5½ year old subjects are significantly lower than the 9½ year old subjects. The other two paired contrasts are not significant.

Table 6 shows the difference between the sample means for the four different sensation levels of the WIPI test, utilizing the Tukey method of multiple comparisons. The mean percentage of correct responses at the 8 dB

TABLE 4.--Results of two way analysis of variance on the WIPI Test for the main effects of age (5½, 7½ and 9½ years) and of sensation level (8, 16, 24 and 32 dB).

Source of Variation	d/f	Mean Squares	F-Ratio
Between Subjects	35		
Age	2	154.78	4.57*
Subjects within age	33	33.89	
Within Subjects	108		
Sensation level (SL)	3	909.15	35.18**
Age x SL	6	21.15	0.82
SL x Subjects within age	99	25.84	
Total	143		

*Significant at $p < .025$.
**Significant at $p < .001$.

TABLE 5.--Difference among means on the WIPI Test: Age main effect.

	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$
$\bar{A}_1$ = 5½ years	0.000	-2.667	-3.419*
$\bar{A}_2$ = 7½ years		0.000	-0.750
$\bar{A}_3$ = 9½ years			0.000

*Significant at the .05 level using Tukey Post Hoc procedures for simultaneous contrasts.

TABLE 6.--Difference among means on the WIPI test: Sensation level main effect.

	$\overline{SL}_1$	$\overline{SL}_2$	$\overline{SL}_3$	$\overline{SL}_4$
$\overline{SL}_1$ (8 dB)	0.000	-8.111*	-10.000*	-11.111*
$\overline{SL}_2$ (16 dB)		0.000	- 1.889	- 3.000
$\overline{SL}_3$ (24 dB)			0.000	- 1.111
$\overline{SL}_4$ (32 dB)				0.000

*Significant at the .05 level using Tukey Post Hoc procedures for simultaneous contrasts.

sensation level is significantly lower than the mean percentage of correct responses made at the three higher sensation levels. No other paired contrasts of means on sensation level are significant.

For purposes of including the data found on the 3½ year old subjects, a separate analysis was done on the four age groups (3½, 5½, 7½ and 9½ years) at three sensation levels (16, 24 and 32 dB). Table 7 summarized the results of this analysis. The F statistic for the mean effects of both age and sensation level were found to be statistically significant. In addition, there was a significant interaction effect. This may be due to the inclusion of the 3½ year old children in this portion of the analysis.

TABLE 7.--Results of two way analysis of variance on the WIPI test for the main effects of age (3½, 5½, 7½ and 9½ years) and of sensation level (16, 24 and 32 dB).

Source of Variation	d/f	Mean Squares	F-Ratio
Between Subjects	47		
Age	3	1074.52	23.12*
Subjects within Age	44	46.48	
Within Subjects	96		
Sensation Level (SL)	2	381.78	14.91*
Age x SL	6	83.85	3.27**
SL x Subjects within Age	88	25.61	
Total	143		

*Significant at $p < .001$.
**Significant at $p < .005$.

Post hoc procedures, the results of which are displayed in Table 8, reveal that the main effect of age is significant only among contrasts including the 3½ year old subjects. Contrasts among the older age groups are not significant.

Table 9 reveals that the main effect of sensation level is significant at a 16 dB sensation level but not at the two higher levels. It will be remembered that the contrasts at the 16 dB sensation level were not significant when the mean scores of children aged 3½ years were not included. Thus it appears that the significant finding at 16 dB SL is attributable to the 3½ year old children.

TABLE 8.--Difference among means on the WIPI Test: Age main effect.

	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$	$\bar{A}_4$
$\bar{A}_1$ (3½ years)	0.000	-9.445*	-10.889*	-11.889*
$\bar{A}_2$ (5½ years)		0.000	- 1.444	- 2.444
$\bar{A}_3$ (7½ years)			0.000	- 1.000
$\bar{A}_4$ (9½ years)				0.000

*Significant at the .05 level using Tukey Post Hoc procedures for simultaneous contrasts.

TABLE 9.--Difference among means on the WIPI Test: Sensation level main effect.

	$\overline{SL}_1$	$\overline{SL}_2$	$\overline{SL}_3$
$\overline{SL}_1$ (16 dB)	0.000	-3.833*	-5.500*
$\overline{SL}_2$ (24 dB)		0.000	-1.667
$\overline{SL}_3$ (32 dB)			0.000

*Significant at the .05 level using the Tukey Post Hoc procedures for simultaneous contrasts.

WIPI Variability of Scores

Further examination of Tables 2 and 3 reveals that the variability of scores, as exhibited in the standard deviations, decreases rather systematically with higher sensation levels. For 5½ year old subjects, for example, the standard deviations on this 25 item test decrease from 8.20 percent at an 8 dB sensation level to 2.61 percent at a 32 dB sensation level.

At a given sensation level, variability also appears to bear an inverse relationship to the age of the subjects. Standard deviations at a 16 dB sensation level were 10.14 percent and 3.17 percent for subjects aged 3½ and 9½ respectively and were 6.26 percent and 1.80 percent for these same ages at a 32 dB sensation level. The variability of the scores at a given sensation level was quite similar between adjacent age groups, except between the 3½ and 5½ year old age levels, where the greatest difference in variability occurred.

WIPI Test-Retest Relationships

Table 10 indicates the differences between mean performance of the subjects from test to retest, in percent.

Differences between the mean scores of the subjects from test to retest appear to be small, differing by more than 1 item, or 4 percent, only at the 8 dB sensation level of the 5½ year old age group. Even there the difference

TABLE 10.--Differences, in percent,* between the mean test and retest scores of 12 subjects from each of the age groups 3½, 5½, 7½ and 9½ years on the WIPI test. Children aged 3½ years were retested only at the 24 dB sensation level.

	Sensation Level			
Age	8 dB	16 dB	24 dB	32 dB
3½	--	--	0.68	--
5½	-4.67	-1.67	0.35	-1.01
7½	-1.01	-2.01	-1.35	-0.32
9½	-1.01	-1.01	-0.32	-0.33

*Negative scores mean that retest scores were higher than test scores.

was only 1.17 items, or 4.67 percent. This is in essential agreement with Ross and Lerman (1970), who found WIPI test-retest scores among hearing impaired children to differ by no more than the equivalent of one word.

As in the Ross and Lerman study, most differences, although small, were in the direction of better scores in the retest session (indicated by negative differences), suggesting the possibility of some slight but consistent learning effect.

Test-retest scores within each age group were so homogeneous that it was felt a test-retest reliability coefficient would not be appropriate. Test-retest reliability is demonstrated by the consistently small mean

differences between test and retest scores at all age levels.

WIPI List and Item Information

It should be pointed out that the pictures of the stimulus items used in this study were the same as those used by Ross and Lerman (1970) in their study with hearing impaired children. However, the pictures were redrawn immediately prior to the publication of the WIPI test. There may well be some variation in the way children respond to the published version of the stimulus items compared to those used in the present investigation and by Ross and Lerman (1970).

Table 11 displays the mean number of items correct and the standard deviation for each WIPI list, with all sensation levels combined, for the test session. Each age group is shown separately.

It can be observed that the mean scores were highly similar from list to list, at all age levels. The maximum difference in mean scores between lists is evidenced between List 2 and List 4 among 3½ year old subjects, and is equal to 2.67 items, or 10.68 percent. In no other age group do the mean scores between test lists differ by more than 1 item, or 4 percent. This generally substantiates the findings of Ross and Lerman (1970) who reported mean list differences to be less than one word among hearing impaired children.

TABLE 11.--The mean (M) number of items correct and the standard deviation (SD) of each WIPI list for the test session (all sensation levels combined) for children aged 3½, 5½, 7½ and 9½ years.

Age		M	SD
3½	List 1	90.24	6.64
	List 2	91.12	7.96
	List 3	83.12	12.92
	List 4	80.44	11.20
5½	List 1	93.68	7.52
	List 2	92.68	6.32
	List 3	93.68	7.52
	List 4	90.00	11.00
7½	List 1	96.32	5.24
	List 2	93.32	6.64
	List 3	96.64	7.96
	List 4	94.32	4.32
9½	List 1	95.32	8.16
	List 2	95.68	3.60
	List 3	96.32	5.24
	List 4	96.00	6.16

Table 12 exhibits the range of difficulty level of items on the WIPI test, with all lists combined. Only errors from the test session are included. It can be observed from this table that 21 of the 100 items of the WIPI test were not missed by any age subject, at any sensation level. Twenty-four words were missed only once. Therefore, the easiest 45 words of the list account for less than 8 percent of the total errors made; indeed, 60 words are responsible for only 17 percent of the total test errors.

TABLE 12.--Item analysis of the WIPI test, all lists combined.

Number of Words	Number of times each word was missed	Column 1 x Column 2	Percentage of the total errors for which these words were responsible	Accumulative percentage of errors for which these words are responsible (easy to difficult).
21	0	0	0	
24	1	24	0.07	0.07
15	2	30	0.09	0.17
8	3	24	0.07	0.24
8	4	32	0.10	0.34
5	5	25	0.08	0.42
3	6	18	0.05	0.48
7	7	49	0.15	0.63
3	8	24	.07	0.68
2	10	20	0.06	0.76
1	14	14	0.04	0.81
1	19	19	0.06	0.87
1	21	21	0.06	0.93
1	22	22	0.07	1.00
$\sum$ =100		$\sum$ =322		

Of the more difficult words, three items of the 100-word WIPI test are responsible for almost 20 percent of the total test errors. Adding the next most frequent error, four words account for nearly one-fourth of the test errors, and six words for almost 30 percent.

Tables 13, 14, 15 and 16 display the number of times each individual word in the four WIPI lists were missed. The total pool of each list is divided into subgroups on the basis of the number of times the words were missed.

Table 13 indicates that the three most difficult words of List 1 (pan, fox, bread) account for 40 percent of the total number of errors made by all subjects on this 25 item word list. When the two next most difficult words of the list (knee, pail) are added, 56 percent of the total number of errors of the list are accounted for. Of the easiest items, 10 of the 25 words are responsible for only 9 percent of the errors made on the word list.

In List 2, four words (bowl, tea, coat, fan) are responsible for 63 percent of the errors. In contrast, the easiest 15 words are responsible for only 6 percent of the errors.

In List 3, six words of the 25 item word list account for nearly 40 percent of the total list errors, while 11 words are responsible for only 7 percent.

In List 4, three words account for almost 40 percent of the list errors, and 11 items account for only 8 percent.

TABLE 13.--Item analysis of the WIPI test: List 1.

Word	Number of times each word was missed	Percentage of the total errors for which these words were responsible.	Accumulative percentage of errors for which these words are responsible (easy to difficult).
bus	0	0	
hat	0		
smoke	0		
train	0		
crib	1		
eye	1		
mouse	1		
neck	1		
shirt	1		
stair	1	0.09	0.09
arm	2		
ball	2		
floor	2		
gun	2		
straw	2		
street	2		
wheel	2	0.21	0.30
chick	3		
school	3		
wing	3	0.14	0.44
knee	5		
pail	5	0.15	0.59
bread	6	0.09	0.68
fox	7	0.11	0.79
pan	14	0.21	1.0

TABLE 14.--Item analysis of the WIPI test: List 2.

Word	Number of times each word was missed	Percentage of the total errors for which these words were responsible	Accumulative percentage of errors for which these words are responsible (easy to difficult).
cake	0	0	
church	0		
clown	0		
dog	0		
door	0		
flag	0		
seal	0		
ship	0		
socks	0		
stick	0		
bear	1		
broom	1		
nail	1		
rug	1		
string	1	0.06	0.06
red	2		
thumb	2	0.05	0.11
barn	3		
desk	3	0.08	0.19
meat	7		
pie	7	0.18	0.37
fan	8	0.10	0.47
coat	10		
tea	10	0.25	0.72
bowl	22	0.28	1.0

TABLE 15.--Item analysis of the WIPI test: List 3.

Word	Number of times each word was missed	Percentage of the total errors for which these words were responsible	Accumulative percentage of errors for which these words are responsible (easy to difficult).
bell	0	0	
chair	0		
cup	0		
snake	0		
car	1		
dirt	1		
dish	1		
feet	1		
sun	1	0.07	0.07
can	2		
key	2		
queen	2	0.08	0.15
bib	3		
nest	3	0.08	0.23
bag	4		
box	4		
coin	4		
fly	4		
saw	4	0.27	0.49
crown	5		
jail	5	0.13	0.63
coke	6	0.08	0.71
moon	7		
thread	7	0.19	0.89
spring	8	0.11	1.00

TABLE 16.--Item analysis of the WIPI test: List 4.

Word	Number of times each word was missed	Percentage of the total errors for which these words were responsible.	Accumulative percentage of errors for which these words are responsible (easy to difficult).
man	0	0	
star	0		
tie	0		
dress	1		
fish	1		
goat	1		
gum	1		
horn	1		
plane	1		
spoon	1		
tail	1	0.08	0.08
bed	2		
bug	2		
frog	2	0.06	0.14
ring	3	0.03	0.17
green	4		
lip	4		
skirt	4	0.12	0.28
black	5	0.05	0.33
pear	6	0.06	0.39
blocks	7		
mouth	7	0.14	0.53
bee	8	0.08	0.61
beet	19	0.19	0.79
bow	21	0.20	1.00

Although comments about the stimulus pictures of the WIPI were not solicited, several were expressed spontaneously. An apparent effort to keep any one item of any matrix from drawing an inappropriate number of responses through sheer attraction of the picture produced some interesting comments. The "queen" is pictured as a rather plain, crowned woman. It was clear that the drab woman did not live up to the children's concept of what a queen should look like. Two children, who had just completed the WIPI, returned to the testing suite with their sister, stating, "We want to show her the ugly queen." A 5½ year old, after pointing correctly to the queen remarked, "I hate that queen. She don't look good. Her necklace and her crown looks good but her face don't look good." A 9½ year old, selecting the same correct stimulus item said, "Look at that ugly girl. That's the ugliest queen I've ever seen." Another 9½ year old said very simply but expressively, "She was ugly."

After the completion of the test, the examiner sometimes pointed to the pictures of certain stimulus items the subject had missed and asked the subject to identify what was in the picture. The picture for the stimulus word "bowl," which was the most frequently missed item of the test, often elicited the response "dish" from the children. The picture for the stimulus word "bow," which was the second most missed item of the test, often elicited the response, "ribbon."

But it was the picture for the stimulus word "beet," third most often missed item of the test, for which the most varied number of responses were recorded, and about which the children appeared to be most confused. Several children looked at all of the pictures after hearing this stimulus item, then either verbalized and pointed to a different stimulus word on the page or appeared to choose the correct picture through a process of elimination. For example, one 5½ year old child said, "Beet? He must have said 'feet,'" then pointed to the picture for the item "feet." Another 5½ year old subject also hesitated after the word "beet," looked at all of the pictures, said "feet" aloud although he had not repeated any of the other stimulus items, and then pointed to the item "feet." Still another 5½ year old subject pointed to the correct picture but remarked, "There wasn't nothing to pick but that one." Another pondered the pictures for some seconds, then pointed to the correct item saying, "Well, I'll say this one."

Following the test sessions, when the examiner pointed to the stimulus item for "beet" and asked what was in the picture, responses included: "cranberries" (three times), "cherries," cookies," and "slices of ham." One 7½ year old child frowned for some length of time as he looked at the picture and finally said, "I don't know. But I know I've seen that before." Four children of those questioned simply stated they did not know what was pictured.

One subject, a 7½ year old girl who was obviously familiar with the item in the picture responded: "Ugh! I don't eat beets."

The item "skirt" was often a troublesome one for the children because included in the same matrix as the picture of the skirt is a picture of a girl in a dress with a pleated skirt. Eight subjects pointed first to the pleated skirt on the girl, although four of these subjects corrected themselves spontaneously.

Summary

The mean responses of children aged 3½, 5½, 7½ and 9½ on the WIPI test were positively correlated with the age of the subject and with the sensation level of the stimulus, with older children and higher sensation levels producing higher mean scores. There was a small difference (approximately one item or less) between test and retest mean scores at any sensation level, for any age group with higher scores on the retest. The mean scores of the four lists were highly similar. Several items of each test list accounted for a relatively large percentage of the errors made on that list.

The PBK-50 Test

The PBK-50 test was constructed by Harriet Haskins, in an attempt to simplify yet maintain the design of the PB-50 lists. This test is composed of words which have been found to be present in the speech of children before

they enter first grade. Haskins constructed four lists but found List 2 to be easier than the others for adult subjects. Subsequently, List 2 has been omitted. What was previously List 3 has become List 2, and what was formerly List 4 has become List 3.

In this study, the 3½ year old subjects were administered one list of the PBK-50 words at a 32 dB sensation level, followed by the same list approximately 15 minutes after the test session as a retest. Subjects aged 5½, 7½, 9½ and 11½ were administered a different list at 16, 24 and 32 dB sensation levels.

Table 17 displays the data obtained with subjects during the test session. Table 18 summarized the same information obtained in the retest session. The means, standard deviations and standard errors of measurement are reported for all of the PBK-50 lists combined and separately for each age group. The mean values reported in these two tables are displayed in graphic form in Figure 2.

PBK-50 Articulation Function

As was noted in the WIPI test, the articulation functions of all age groups are similar in configuration, showing little change between test and retest. Again, it is observed that although the difference in mean scores between adjacent age groups is slight, that difference is positively related to age. The 11½ year old subjects

TABLE 17.--Mean (M), standard deviation (SD) and standard error of measurement (Se) in percent correct of 12 normal hearing children from each of the age groups 3½, 5½, 7½, 9½ and 11½ on the PBK-50 test during the test session.

Age	Sensation Level in dB	M	SD	Se
3½	16	--	--	--
	24	--	--	--
	32	71.67	15.75	5.41
5½	16	85.83	4.86	3.87
	24	92.83	3.66	3.17
	32	95.83	2.48	2.33
7½	16	89.50	4.68	2.76
	24	95.00	2.36	2.70
	32	97.50	1.51	1.12
9½	16	89.17	4.13	3.35
	24	95.83	2.89	2.33
	32	98.50	1.24	1.04
11½	16	89.17	7.51	2.46
	24	97.00	2.89	1.23
	32	98.83	1.80	0.83

TABLE 18.--Mean (M) and standard deviation (SD) in percent correct of 12 normal hearing children from each of the age groups 3½, 5½, 7½, 9½ and 11½ on the PBK-50 test during the retest session.

Age	Sensation Level in dB	M	SD
3½	16	--	--
	24	--	--
	32	69.50	14.48
5½	16	86.33	4.89
	24	92.50	4.36
	32	95.67	3.49
7½	16	89.33	3.94
	24	93.17	4.39
	32	97.67	1.67
9½	16	89.17	6.06
	24	96.33	3.17
	32	98.33	1.67
11½	16	89.67	6.87
	24	97.33	3.65
	32	99.33	1.30

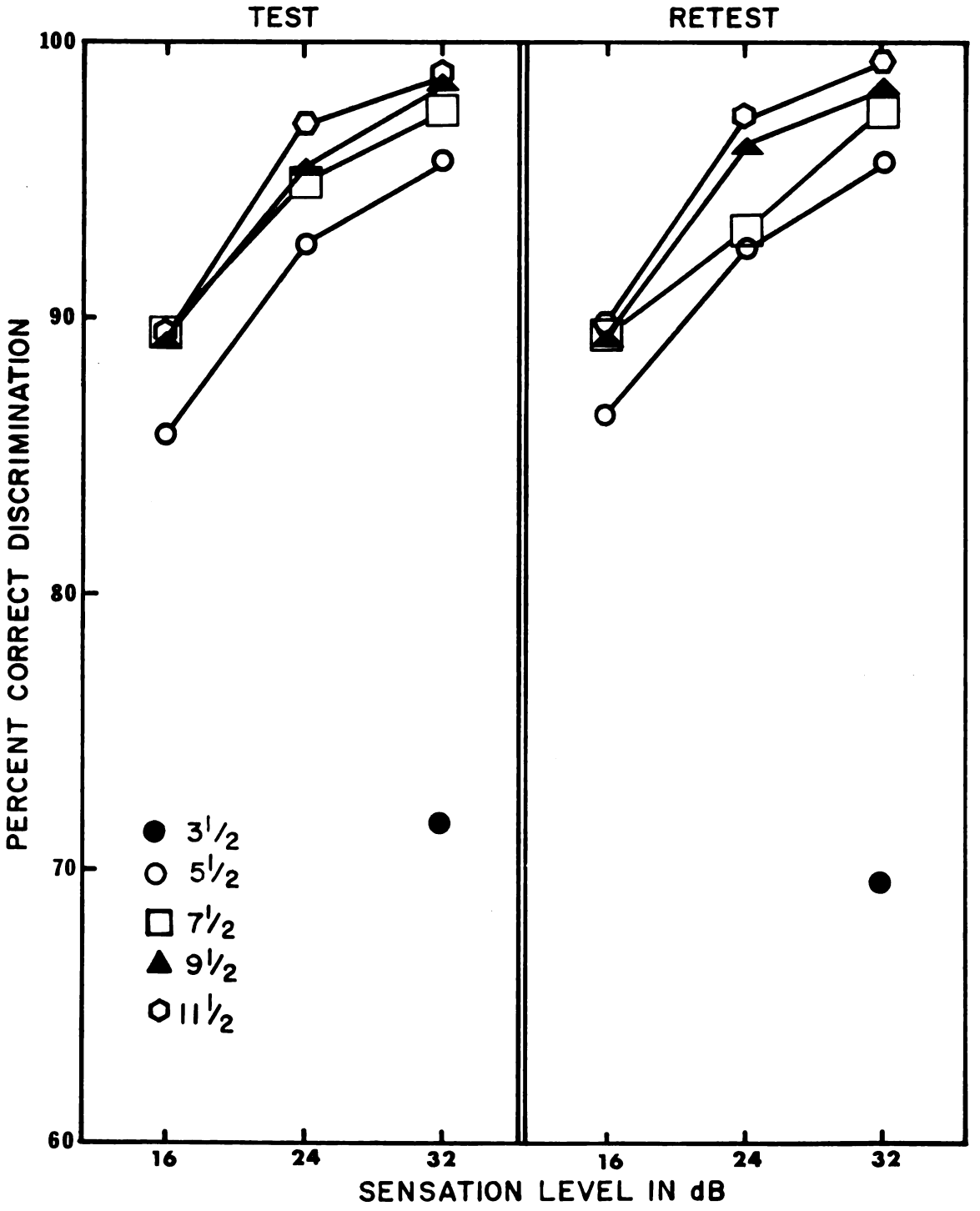

Figure 2.--The mean discrimination scores of 12 normal hearing children from each of the age groups 3½, 5½, 7½, 9½ and 11½ on the PBK-50 during test and retest sessions.

showed slightly better scores than the 9½ year old group, the 9½ year olds slightly surpassing the 7½ year olds, and the 7½ year olds scoring higher than subjects aged 5½. Subjects aged 3½ performed substantially poorer than the 5½ year old subjects at the 32 dB sensation level, the only sensation level at which they were tested.

The results of analysis of variance are summarized in Table 19. The F statistics for the mean effects of both age and sensation level were significant. The interaction effect for this portion of the study was not significant.

Post hoc procedures, using the Tukey method of multiple comparisons, were used to indicate differences between sample means. Table 20 indicates that the mean scores of 5½ year old subjects were signficiantly lower than the scores of children aged 9½ and 11½ years. Other contrasts were not significant.

Table 21 displays the same information for the sensation levels used in the study. Here all possible paired contrasts between the sample means were significant, with highest mean scores obtained at the highest sensation levels.

The articulation functions evidenced in Figure 2 are curvilinear, with progressive increases in stimulus strength accompanied by smaller increments in mean scores. The curve appears to be nearing asymptote at a 32 dB sensation level for subjects aged 7½ and 9½.

TABLE 19.--Results of two way analysis of variance on the PBK-50 test for the main effects of age (5½, 7½, 9½ and 11½ years) and of sensation level (16, 24 and 32 dB).

Source of Variation	d/f	Mean Square	F-Ratio
Between Subjects:	47		
Age	3	87.00	4.49*
Subjects within Age	44	19.38	
Within Subjects:	96		
Sensation Level (SL)	2	1099.00	97.52**
Age x SL	6	3.89	0.35
SL x Subjects within age	88	11.27	
Total	143		

*Significant at $p < .025$.
**Significant at $p < .001$.

TABLE 20.--Difference among means on the PBK-50 test: Age mean effect.

	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$	$\bar{A}_4$
$\bar{A}_1$ (5½ years)	0.000	-2.500	-3.000*	-3.500*
$\bar{A}_2$ (7½ years)		0.000	-0.500	-1.000
$\bar{A}_3$ (9½ years)			0.000	-0.500
$\bar{A}_4$ (11½ years)				0.000

*Significant at the .05 level using the Tukey Post Hoc procedures for simultaneous contrasts.

TABLE 21.--Difference among means on the PBK-50 test: Sensation level main effect.

	$\overline{SL}_1$	$\overline{SL}_2$	$\overline{SL}_3$
$\overline{SL}_1$ (16 dB)	0.000	-6.700*	-9.250*
$\overline{SL}_2$ (24 dB)		0.000	-2.500*
$\overline{SL}_3$ (32 dB)			0.000

*Significant at the .05 level using the Tukey Post Hoc procedures for simultaneous contrasts.

PBK-50 Variability of Scores

Tables 17 and 18 indicate that as the sensation level increases, the variability of scores, as indicated by the standard deviations within age group, systematically decreases. In contrast to the WIPI test, variability does not appear to be inversely related to the age of the subjects at a given sensation level. At the 16 dB sensation level, for example, 11½ year old subjects displayed greater variability than either the 9½ or 7½ year old subjects, although the mean scores for the three age groups were very similar. At a 24 dB sensation level, the 7½ year old subjects had a smaller standard deviation than their 9½ and 11½ year old counterparts. Finally, at the 32 dB sensation level, the 11½ year old subjects scores showed more variability than the 7½ and 9½ year olds. By far the most

variability in test scores was found for the 3½ year old subjects, with a standard deviation of 15.75 percent at a 32 dB sensation level.

Variability of test and retest scores was generally similar and was not consistently lower for either the test or retest sessions.

PBK-50 Test-Retest Relationships

Table 22 indicates the absolute difference between the mean performances of subjects on test and retest.

It can be observed that only with the 3½ year old subjects did the mean score between test and retest differ by more than one item, or 2 percent. Differences between the mean scores are consistently small and do not appear to bear a systematic relationship to either the age of the subjects or to the sensation level. The direction of the change in mean scores is equally divided between better scores for retest (as evidenced in negative differences) and better scores for the test session. Mean scores of the 11½ year old subjects, however, proved consistently although only slightly higher for the retest session.

Scores for an age group were so homogeneous that it was felt a test-retest reliability coefficient would not be appropriate. However, the consistently small difference in mean test and retest scores indicate that the test is a reliable one at this age level.

TABLE 22.--Differences, in percent,* between the mean test and retest scores of 12 subjects from each of the age groups 3½, 5½, 7½, 9½ and 11½ years on the PBK-50 test. Children aged 3½ years were tested only at the 32 dB sensation level.

	Sensation Level		
Age	16 dB	24 dB	32 dB
3½	--	--	2.17
5½	-0.50	0.33	0.16
7½	0.17	1.83	-0.17
9½	0.00	-0.5	0.17
11½	-0.5	-0.33	-0.5

*Negative scores mean that retest scores were higher than test scores.

PBK-50 List Information

Table 23 indicates the mean number of items correct and the standard deviation for each of the PBK-50 lists, combining all sensation levels, for each age group separately.

At the 3½ year level, a difference of 5 items or 10 percent appears between List 2 and List 3. This difference is not, however, substantiated by any of the other age groups. Differences between the mean list scores of the other age groups do not vary by more than 2 items, or 4 percent. This appears to indicate that the three PBK-50 lists are of similar difficulty.

TABLE 23.--The mean (M) number of items correct and the standard deviation (SD) of each PBK-50 list for the test session, with all sensation levels combined, for children aged 3½, 5½, 7½, 9½ and 11½ years.

Age		M	SD
3½	List 1	71.00	14.46
	List 2	77.00	23.24
	List 3	67.00	9.58
5½	List 1	91.50	5.86
	List 2	90.00	6.20
	List 3	93.00	4.78
7½	List 1	94.16	5.00
	List 2	94.50	3.62
	List 3	93.34	5.20
9½	List 1	93.66	6.14
	List 2	94.84	5.36
	List 3	95.00	3.14
11½	List 1	93.16	8.38
	List 2	95.00	5.62
	List 3	96.84	4.04

Summary

The mean responses of children aged 5½, 7½, 9½ and 11½ years on the PBK-50 test were highly similar in configuration, with the mean scores of the 3½ and 5½ year old subjects significantly lower than those of children aged 9½ and 11½. The variability of scores, as indicated by the standard deviations within age groups, bears an inverse relationship to the sensation level of the stimulus items. Subjects aged 3½ showed substantially more variability in test scores than the older age groups. Differences between

test and retest mean scores were consistently small. Differences between the mean scores of the three lists were also small, indicating that the lists are essentially equivalent.

N.U. Auditory Test No. 6

The N.U. Auditory Test No. 6, it will be remembered, is composed of four phonemically equivalent CNC lists. The test was standardized with a group of normal hearing and a group of sensorineural hearing impaired adult subjects (Tillman and Carhart, 1966).

In this study twelve children from each of the age groups 7½, 9½ and 11½ years were administered one list of the N.U. Auditory Test No. 6 at 8, 16, 24 and 32 dB sensation levels. This was followed by a retest, using the same lists at the same sensation levels, approximately 15 minutes after the test session.

Table 24 and Table 25 present data obtained for the N.U. Auditory Test No. 6 during the test and retest sessions. The range, mean, standard deviation and standard error of measurement are indicated for each age group, at every sensation level tested. The mean values of each sensation level are displayed in graphic form in Figure 3.

N.U. Auditory Test No. 6 Articulation Function

Observation of Table 23 reveals that the mean score differences between age groups in the test session were relatively small. Only at the 8 dB sensation level of the

TABLE 24.--The mean (M), standard deviation (SD) and standard error of measurement (Se) in percent correct of 12 normal hearing children from each of the age groups 7½, 9½ and 11½ on the N.U. Auditory Test No. 6 during the test session.

Age	Sensation Level in dB	M	SD	Se
7½	8	56.33	13.15	6.56
	16	81.33	5.00	3.06
	24	88.83	5.36	3.17
	32	92.83	3.66	2.75
9½	8	62.00	5.85	4.55
	16	79.83	6.35	6.73
	24	89.33	4.92	2.62
	32	93.50	2.97	2.55
11½	8	62.67	6.84	6.60
	16	82.33	6.37	4.15
	24	88.83	6.90	2.97
	32	96.50	2.71	1.31

TABLE 25.--The mean (M) and standard deviation (SD) in percent correct of 12 normal hearing children from each of the age groups 7½, 9½ and 11½ on the N.U. Auditory Test No. 6 during the retest session.

Age	Sensation Level in dB	M	SD
7½	8	51.83	11.42
	16	78.17	6.63
	24	88.17	6.24
	32	92.50	3.63
9½	8	61.50	10.59
	16	80.67	7.78
	24	91.50	4.68
	32	95.33	3.75
11½	8	67.67	8.22
	16	84.17	6.58
	24	92.33	4.74
	32	96.00	2.09

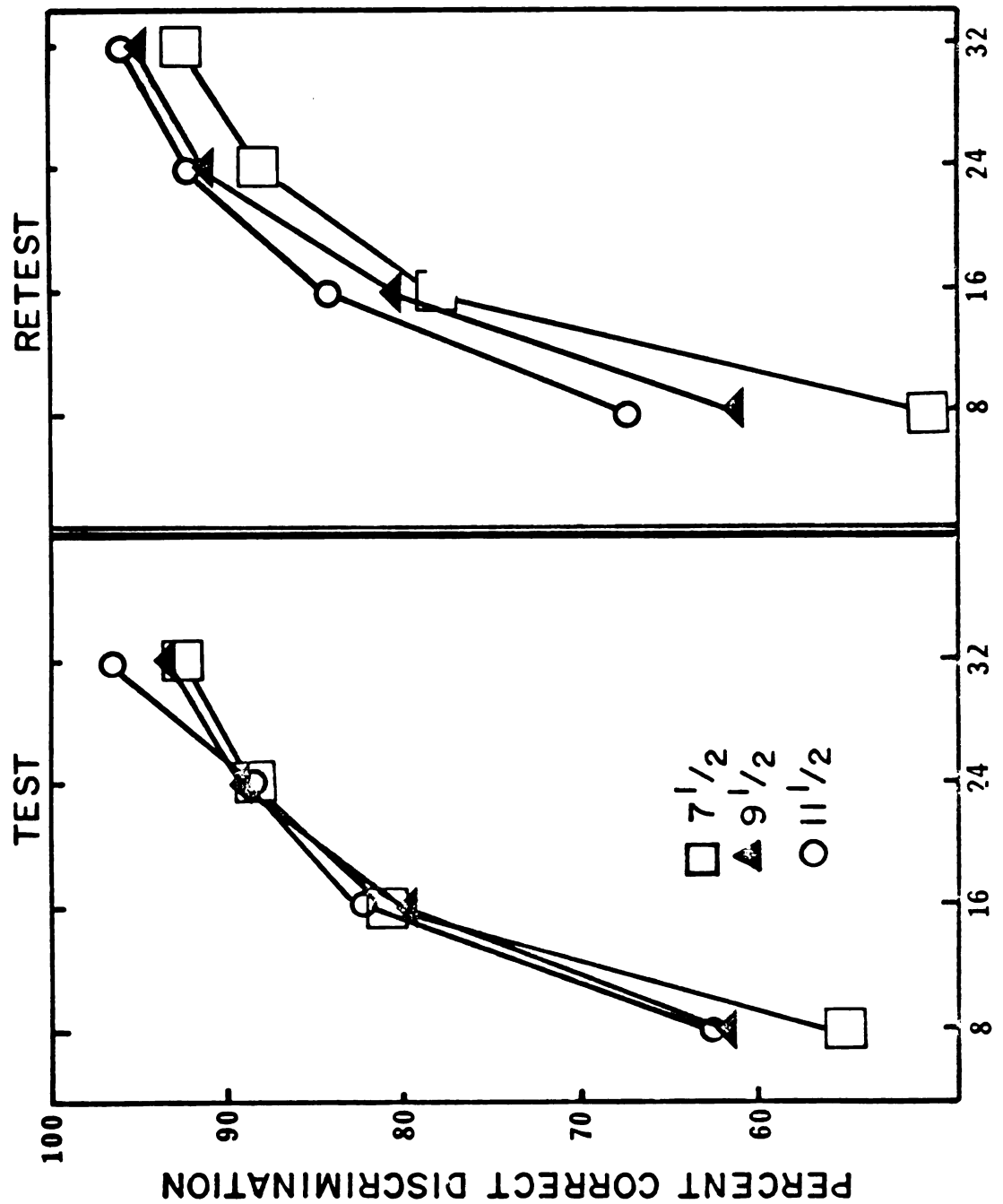

Figure 3.--The mean discrimination scores of 12 normal hearing children from each of the age groups 7½, 9½, and 11½ on the N.U. Auditory Test No. 6 during test and retest sessions.

7½ and 9½ year old age groups did the difference in means between any of the three age groups exceed two items, or 4 percent, at a given sensation level. Observation of the retest scores in Table 25 reveals more spread among the mean scores of the subjects at various ages than in the test session. However, these differences are confined to the lower sensation levels. On retest, a difference of 7.9 items, or nearly 16 percent, existed between the 7½ and 11½ year old subjects at an 8 dB sensation level, with the 11½ year old subjects having higher scores. The difference between these age groups for the test session was only 3.16 items, or 6.32 percent. At this same sensation level the 11½ year old group differed from their adjacent age group, the 9½ year olds, by over 6%, whereas in the test session there was less than 1 percent difference between the two age groups. At the 24 and 32 dB sensation levels, however, differences between adjacent age groups were less than 2 items, or 4 percent.

Analysis of variance, summarized in Table 26, reveals that only the F statistic for the sensation level main effect was significant. Post hoc procedures, shown in Table 27, revealed that the paired contrasts between the mean scores obtained at every sensation level were significantly different from each other, with the highest sensation level exhibiting the highest mean scores.

Observation of Figure 3 reveals a similarity of articulation configurations among the age groups and

TABLE 26.--Results of two way analysis of variance on the N.U. Auditory Test No. 6 for the main effects of age (7½, 9½ and 11½ years) and of sensation level (8, 16, 24 and 32 dB).

Source of Variation	d/f	Mean Squares	F-Ratio
Between Subjects	35		
Age	2	90.78	1.32
Subjects within age	33	69.00	
Within Subjects	108		
Sensation Level (SL)	3	8007.44	254.65*
Age x SL	6	40.11	
SL x Subjects within age	99	31.44	

*Significant at $p < .001$.

TABLE 27.--Difference among means on the N.U. Auditory Test No. 6 Test: Sensation level main effect.

	$\overline{SL}_1$	$\overline{SL}_2$	$\overline{SL}_3$	$\overline{SL}_4$
$\overline{SL}_1$	0.000	-20.834*	-28.667*	-33.945*
$\overline{SL}_2$		0.000	- 7.833*	-13.111*
$\overline{SL}_3$			0.000	- 5.278*
$\overline{SL}_4$				0.000

*Significant at the .05 level using Tukey Post Hoc procedures for simultaneous contrasts.

between test and retest. Although there is some interweaving, there is a general trend for the mean scores to order themselves in relationship to the age of the subjects, particularly in retest. Up to 24 dB sensation level there is a sharp increase in scores as a function of increased intensity. Between 24 and 32 dB sensation level, however, there is less change in scores as a function of intensity, and the curve appears to be reaching asymptote.

N.U. Auditory Test No. 6
Variability of Scores

Variability of test scores at the lower sensation levels does not appear to bear a consistent relationship either to age level or to sensation level in the test session. Consistency here appears to occur only at the 32 dB sensation level, where the standard deviation was smaller for all age groups than in the lower sensation levels. Standard deviations of the retest scores became progressively lower with added intensity at every age level.

N.U. Auditory Test No. 6
Test-Retest Relationships

Table 28 indicates the difference in percent between the test-retest mean scores of subjects at every sensation level.

At the 8 dB sensation level, the difference between the mean test and retest scores of the 7½ and 11½ year old subjects were 4.50 percent and -5.00 percent respectively.

TABLE 28.--Differences, in percent,* between the mean test and retest scores of 12 subjects from each of the age groups 7½, 9½ and 11½ years on the N.U. Auditory Test No. 6.

	Sensation Level			
Age	8 dB	16 dB	24 dB	32 dB
7½	4.50	3.16	0.66	0.33
9½	0.50	-0.84	-2.17	-1.83
11½	-5.00	-1.84	-3.5	0.50

*Negative scores mean that retest scores were higher than test scores.

Although these differences are larger than for the other age groups and sensation levels, they are not consistent in direction, the 7½ year olds having higher mean scores for the test session (as indicated by the positive difference) and the 11½ year old subjects having larger mean scores for the retest session (as indicated by negative differences). In fact, of the 12 observations of mean differences, six showed better scores in test, and six showed improved scores in retest. Both the 9½ and 11½ year old subjects, however, demonstrated improved mean scores at three sensation levels out of four, indicating a possible learning factor in the two older age groups.

Except for the 8 dB sensation level, differences between test and retest mean scores varied by less than 2 items, or 4 percent.

N.U. Auditory Test No. 6 List Information

Table 29 discloses the mean scores and standard deviations of every age level on each list of the N.U. Auditory Test No. 6. Again, the mean scores between lists were highly similar. This appears to substantiate the findings of Tillman and Carhart (1966) and Rintelmann and Jetty (1968), who reported that the lists are essentially equivalent.

Summary

The mean scores of children aged 7½, 9½ and 11½ on the N.U. Auditory Test No. 6 were highly similar. Differences between mean test and retest scores were generally small. The four lists of the test are essentially equivalent.

TABLE 29.--The mean (M) number of items correct and the standard deviation (SD) of each N.U. Auditory Test No. 6 list for the test session, with all sensation levels combined, for children aged 7½, 9½ and 11½ years.

Age		M	SD
7½	List 1	76.16	22.56
	List 2	79.50	13.92
	List 3	80.00	13.34
	List 4	83.66	14.36
9½	List 1	81.64	15.40
	List 2	80.66	14.56
	List 3	81.84	10.18
	List 4	80.34	14.04
11½	List 1	80.50	17.84
	List 2	80.34	13.58
	List 3	83.00	11.14
	List 4	86.50	13.22

Discussion

The following portion of this chapter discusses the three tests used in this study in terms of their clinical application with children at the several age levels investigated.

Children Aged 3½ Years

Figure 4 reveals the results of the WIPI and the PBK-50 tests administered to subjects aged 3½ years during the test session. The numbers in parentheses are the standard deviations in percent.

For the assessment of speech discrimination performance of children in this age group, the 25 item WIPI test appears to be a particularly suitable instrument. At a 32 dB sensation level, the mean score of normal hearing children in this age group was 91.68 percent with a standard deviation of 6.26 percent. At a 24 dB sensation level the mean was almost as high (88.08%), but at this sensation level the standard deviation increased to 10.85 percent.

On the PBK-50 test, the mean score of 3½ year old children with normal hearing was only 35.83 items out of 50, or 71.66 percent. The variability of the scores of this test was much greater than on the WIPI, as indicated by the standard deviation of 15.75 percent even within the rather homogeneous population used in this study. This suggests that factors other than hearing itself may be

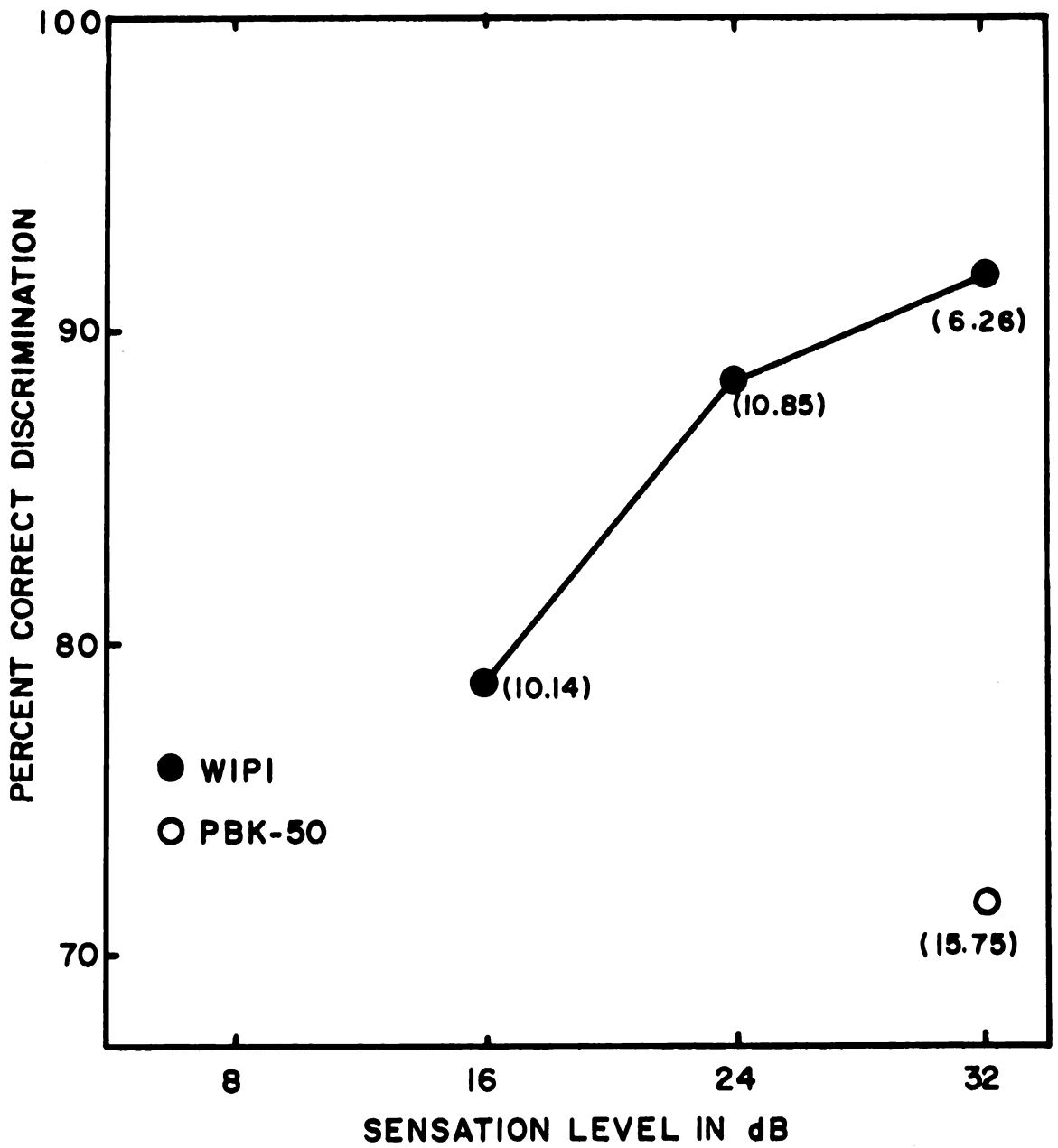

Figure 4.--Mean discrimination scores and standard deviations (in parentheses) of 12 normal hearing children aged 3½ years on the WIPI and the PBK-50.

influencing test scores. Such factors may include speech problems, short attention span, the open set as opposed to the closed set, and vocabulary variables.

For children aged 3½ years, the WIPI test administered at a 32 dB sensation level appears to have the double advantage of allowing normal hearing children to obtain good scores and of having less variability among the scores within the particular population. A further advantage of the WIPI test is the possibility that the closed set of the test may allow scores to be easily obtained even from children who, because of personality factors or speech inadequacies, find it difficult to respond to more conventional speech audiometry.

The tester should be aware that there are several test items which appear to cause inappropriate responses either because of the characteristics of the picture or because of children's unfamiliarity with the test item. These test items include bowl (list 2), bow (list 4), beet (list 4) and skirt (list 4). Lists 1 and 3 appear to be the most trouble-free in respect to individual test items. However, because the picture stimulus items were changed between the time of this study and the time of the publication of the test, generalizations are difficult.

Children Aged 5½ Years

Figure 5 reveals the mean articulation scores and standard deviations (in percent) of 5½ year old children, for the test sessions of the WIPI and the PBK-50.

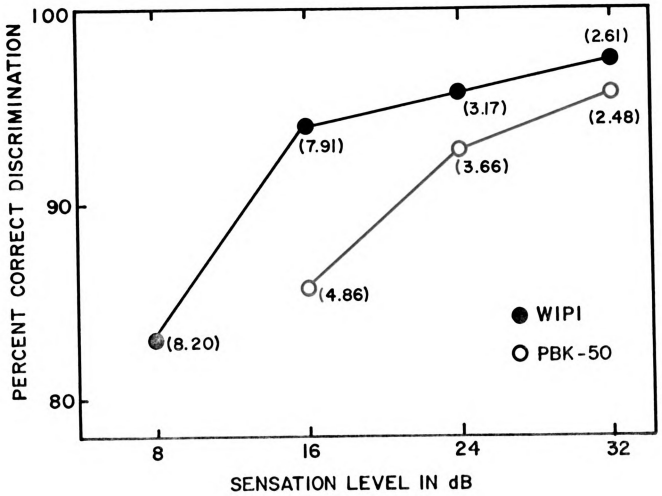

Figure 5.--Mean discrimination scores and standard deviations (in parentheses) of 12 normal hearing children aged 5½ years on the WIPI and the PBK-50.

For children aged 5½ years it also appears that the WIPI is an appropriate clinical tool. The mean score of normal hearing children at a 32 dB sensation level was over 97 percent, with an observed standard deviation of 2.6 percent. The small standard deviation and the small difference in test-retest scores mentioned earlier suggest that the test is a reliable one at this age level.

On the PBK-50 test, the mean score for 5½ year old children at a 32 dB sensation level was almost 96 percent. This test also appears to be appropriate for children aged 5½ years, provided the 5½ year old children have normal speech and language development. The mean percentage scores of 5½ year old children with normal hearing on the WIPI were very similar to those of the same children on the PBK-50 test, at both the 24 and 32 dB sensation levels. Percentage scores between the two tests at these sensation levels, in fact, vary by less than 4 percent. Taking into consideration the formats of the tests, the WIPI test might be preferred on the grounds that it is shorter, it is not limited by speech and language dissimilarities among subjects of this age group, and simply because it is an intrinsically more interesting task for children of this age group.

Children Aged 7½ and 9½ Years

The test results of children aged 7½ and 9½ years will be discussed together, because of the similarities in test results of the two age groups.

Figure 6 reveals the mean articulation scores and the standard deviations (in percent) for children aged 7½ years on the test session of the WIPI, the PBK-50, and the N.U. Auditory Test No. 6. Figure 7 reveals the same information for children aged 9½ years.

It can be observed that both of these age groups achieved high mean scores on the WIPI test at both the 24 and 32 dB sensation levels, with little variability of scores as exhibited by a 2.61 percent standard deviation of the 7½ year old group at both the 24 and 32 dB sensation levels, and 1.97 percent and 1.80 percent standard deviations by the 9½ year old group at 24 and 32 dB sensation levels respectively.

Of the two conventional speech discrimination tests used in this study, the PBK-50 is the easier for these age groups as evidenced by better mean articulation scores at every sensation level, by less variability of test scores, and by less difference in test-retest mean scores. Subjects aged 7½ years attained a mean score of 97.50 percent at the 32 dB sensation level on the PBK-50 with a standard deviation of 1.51 percent. The 9½ year old subjects, at the same sensation level, had a mean score of 98.50 percent with a standard deviation of 1.24 percent. On the N.U. Auditory Test No. 6, at the 32 dB sensation level, 7½ year old subjects obtained a mean score of 92.84 percent with a standard deviation of 3.66 percent, and 9½ year old subjects

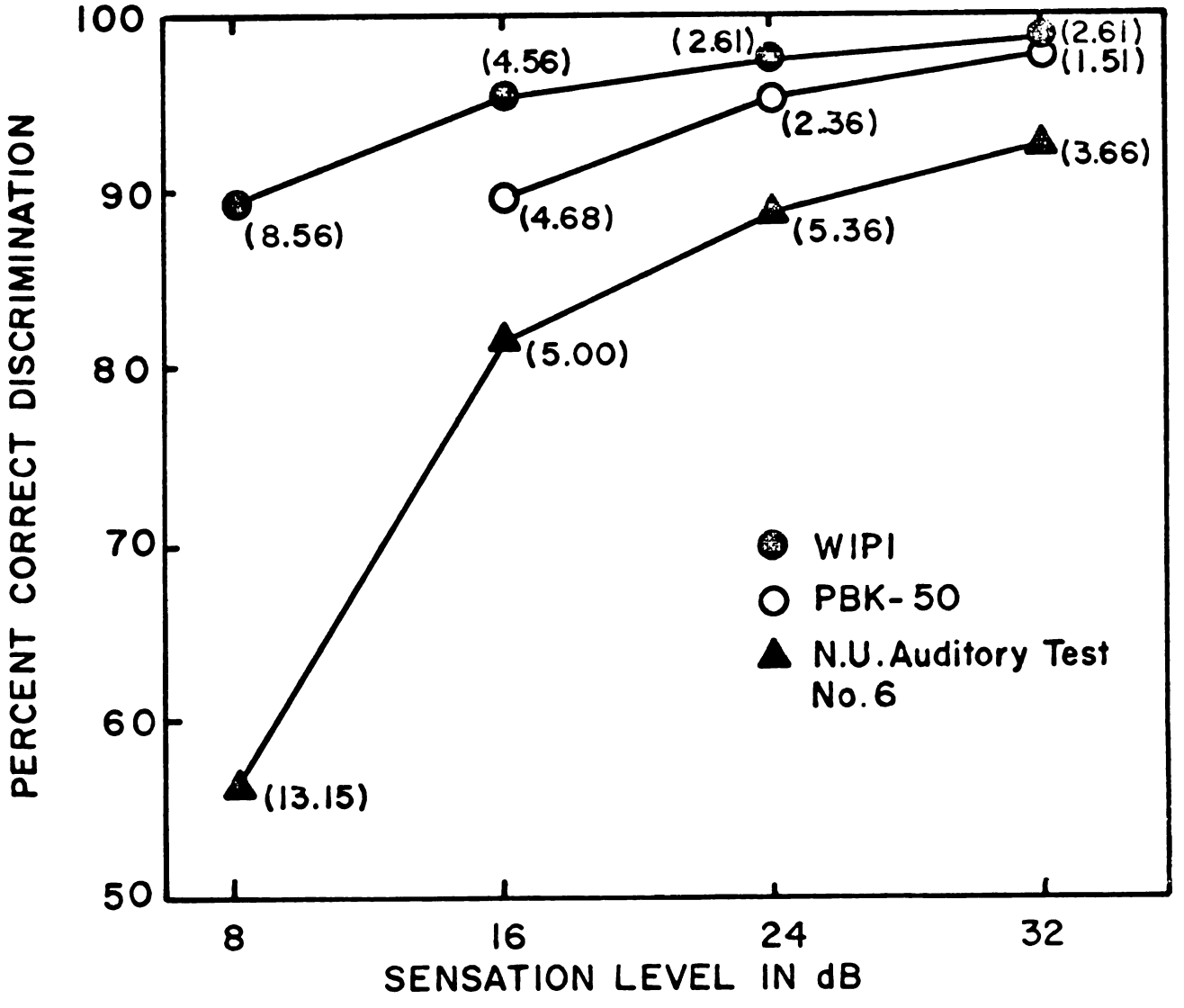

Figure 6.--Mean discrimination scores and standard deviations (in parentheses) of 12 normal hearing children aged 7½ years on the WIPI, the PBK-50, and the N.U. Auditory Test No. 6.

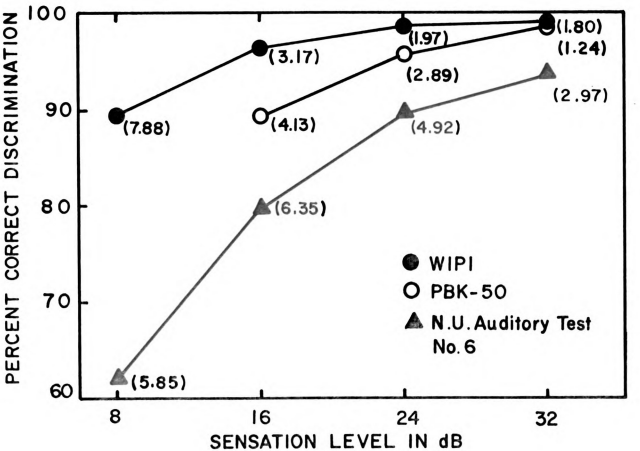

Figure 7.--Mean discrimination scores and standard deviations (in parentheses) of 12 normal hearing children aged 9½ years on the WIPI, the PBK-50, and the N.U. Auditory Test No. 6.

a mean score of 93.50 percent with a standard deviation of 2.97 percent.

As pointed out by Carhart (1965), the purpose for which the discrimination testing is being done should dictate the selection of the specific test to be administered. If the purpose of the test is to determine whether the hearing of the child falls within normal limits, the WIPI and the PBK-50 appear to be the tests of choice, allowing high scores among normal hearing children of these age groups, with little dispersion of test scores.

However, if testing is to be done for purposes of evaluating a hearing aid or of determining differences in auditory discrimination ability between ears, it is suggested that the more difficult N.U. Auditory Test No. 6 be used, where a wider dispersion of speech discrimination scores can be obtained. In repeatedly administering the test to the same child, speech and language development is held constant; thus, differences in test scores can be assumed to be due to differences in hearing as measured by an appropriate speech discrimination test. On an easier task, with a low test ceiling, little change in discrimination scores may be found even though a real difference in hearing is present.

Children Aged 11½ Years

Figure 8 displays the mean articulation scores and the standard deviations (in percent) for children aged 11½

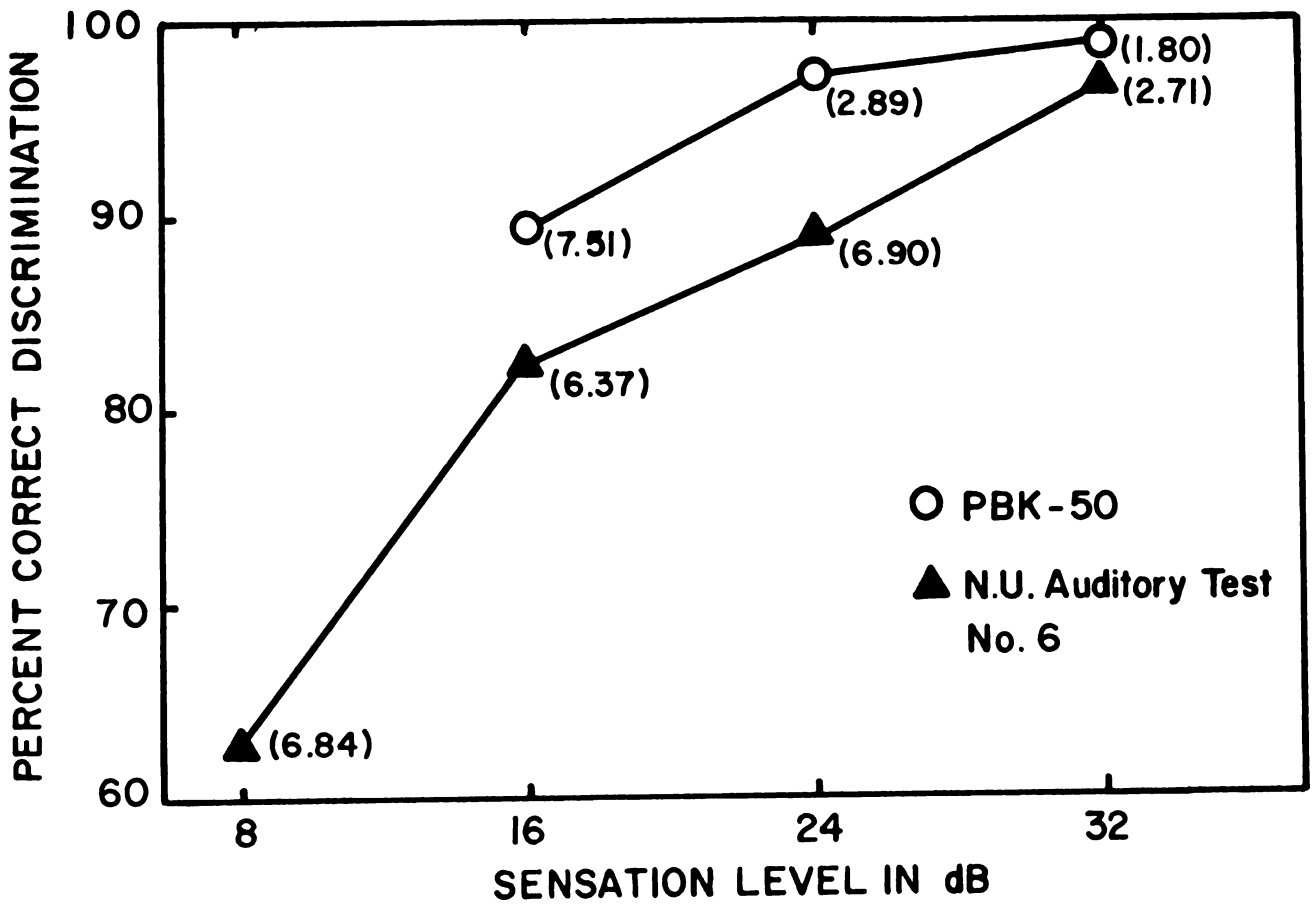

Figure 8.--Mean discrimination scores and standard deviations (in parentheses) of 12 normal hearing children aged 11½ years on the PBK-50 and the N.U. Auditory Test No. 6.

years on the PBK-50 and the N.U. Auditory Test No. 6. The WIPI test was not administered to this age group.

Observation of the articulation functions of the two tests administered to children in this age group again reveals that the N.U. Auditory Test No. 6 is more difficult, although the difference in mean scores between the two tests was only 2.33 percent at a 32 dB sensation level.

Performance of the 11½ year old children on the N.U. Auditory Test No. 6 closely approximates the findings of Tillman and Carhart (1966) with normal hearing adults at the 32 dB sensation level. The mean score of the 11½ year old subjects at this sensation level was 96.50 percent, compared to means ranging between 98.60 percent and 99.60 percent on the four lists for the normal hearing adults of that study. Differences in the mean scores between the 11½ year old subjects and adults were somewhat greater in the lower sensation levels.

Again, as already stated regarding 7½ and 9½ year old subjects, it would appear that the purpose of the discrimination testing dictate the particular test to be administered to children in the 11½ year age group. For purposes of determining whether the child has normal hearing, the PBK-50 appears to be the preferred test, whereas the N.U. Auditory Test No. 6 is more appropriate where a more difficult test is required.

Summary

Of the two tests administered to children aged 3½ years, the multiple choice, WIPI test appears to be the "test of choice" with higher and less variable mean scores than those obtained with the PBK-50 test.

For children aged 5½ years, the PBK-50 appears to be an appropriate tool, providing the child has good speech and normal language development.

For children aged 7½, 9½ and 11½ the N.U. Auditory Test No. 6 is more difficult than the PBK-50 test. The purpose for which the discrimination testing is being done should dictate the selection of the specific test to be administered to children of these age groups.

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this study was to investigate the performance of a group of normal hearing children on three types of speech discrimination materials: (1) multiple-choice discrimination tests requiring no verbal response on the part of the child, (2) conventional discrimination tests which have been modified for use with children, and (3) standardized discrimination tests for adults. Tests used to represent these three types of speech materials were the Word Intelligibility by Picture Identification Test (WIPI), the Phonetically Balanced Kindergarten Word Lists (PBK-50), and the Northwestern University Auditory Test No. 6 (N.U. Auditory Test No. 6), respectively.

Summary

Tape-recorded versions of three speech discrimination tests, the WIPI, the PBK-50, and the N.U. Auditory Test No. 6 were administered to sixty normal hearing children, 12 from each of the following age groups: 3½, 5½, 7½, 9½ and 11½ years. Because of the age range of the

sub-groups tested, the number of tests and sensation levels administered varied by age group. All subjects received a retest approximately 15 minutes following the test session. The three speech discrimination tests were administered on separate days.

In determining the articulation function for each of the three discrimination tests, every successive presentation was given at a higher intensity with a different list of the test given at each sensation level tested. For the retest, an alternate randomization of the same list of words was presented at each sensation level.

Representing multiple-choice picture discrimination tests requiring no verbal response on the part of the child, the WIPI test was selected. Twelve children with normal hearing from each of the age groups 3½, 5½, 7½ and 9½ years were administered the WIPI test in this study. This list was administered at 8, 16, 24 and 32 dB sensation levels for all subjects except those aged 3½ years, who were given the test at all but the 8 dB sensation level.

Articulation functions yielded by the children of the different age levels on the WIPI test were highly similar to one another in configuration, with little change between test and retest. Analysis of variance confirmed that the effects associated with both age and sensation level were statistically significant. Although the differences in the mean percent correct responses between adjacent age groups were small, they were positively

correlated with the age of the subjects. Little difference was evidenced between mean test and retest scores. Variability of scores, as exhibited by the standard deviations, decreased with higher sensation levels and bore an inverse relationship to the age of the subjects at a given sensation level. Mean scores of the four lists were similar. Three items of the 100-word WIPI test were responsible for almost 20 percent of the total test errors.

The PBK-50 test was selected to represent conventional discrimination tests which have been modified for use with children. The 3½ year old subjects were administered one list of the PBK-50 words at a 32 dB sensation level, followed by the same list approximately 15 minutes after the test session. Subjects aged 5½, 7½, 9½ and 11½ were administered a different list at 16, 24 and 32 dB sensation levels.

The articulation functions of children aged 5½, 7½, 9½ and 11½ years on the PBK-50 were highly similar in configuration. The mean scores of subjects aged 3½ and 5½ years were significantly lower than those of children aged 9½ and 11½ years. Speech discrimination scores at every sensation level were found to be significantly different from scores at adjacent sensation levels, with the highest mean scores obtained at the highest sensation levels. The variability of scores, as indicated by the standard deviations within age groups, was inversely related to the sensation level of the test items. The differences

in mean test-retest scores were consistently small. The mean scores of the three test lists indicated that the three lists were essentially equivalent.

To represent standardized discrimination tests for adults, the N.U. Auditory Test No. 6 was selected and administered to subjects aged 7½, 9½ and 11½ years, at 16, 24 and 32 dB sensation level. The mean scores of the subjects of the three age groups were highly similar. Variability of the test scores at a 32 dB sensation level was smaller than at lower sensation levels. Differences between mean test-retest scores were small; however, the two older age groups evidenced slightly higher mean scores for the retest on three of the four sensation levels tested, indicating a possible learning factor. Mean scores on the four lists were similar.

Conclusions

Within the limits of the present study, the following conclusions appear warranted.

1. The speech discrimination scores obtained for a child at a given sensation level depends, in part, upon the particular discrimination test selected for use with that child.
2. For the assessment of speech discrimination of 3½ year old children, the 25 item multiple-choice WIPI test is a suitable instrument. At a 32 dB sensation level, the score of a normal hearing child can be expected to be close to 92 percent on this test.

On the PBK-50 test at a 32 dB sensation level, the score of a 3½ year old child with normal speech and language development can be expected to be only about 72 percent, suggesting that factors other than hearing itself may be influencing test scores.

3. For children aged 5½ years, the WIPI is also an appropriate clinical tool. The score of a normal hearing child of this age at a 32 dB sensation level can be expected to be close to 97 percent.

The PBK-50 test is appropriate for children in this age group, provided the children have normal speech and language development. The score of a 5½ year old normal hearing child at a 32 dB sensation level on the PBK-50 test can be expected to be close to 96 percent.

4. Children aged 7½ years obtain scores very similar to children aged 9½ years on the WIPI test and to children aged 9½ and 11½ years on the PBK-50 test and the N.U. Auditory Test No. 6. At a 32 dB sensation level, they can be expected to obtain scores close to 98 percent on the WIPI and PBK-50 lists and 93 percent on the N.U. Auditory Test No. 6.
5. For children aged 7½, 9½ and 11½ years, the N.U. Auditory Test No. 6 is more difficult than the PBK-50, as evidenced by steeper articulation functions, lower discrimination scores at all sensationlevels, and greater variability of test scores.

6. For all age groups on all three tests, the variability of test scores as exhibited by the standard deviation is smaller at a 32 dB sensation level than at the lower sensation levels. In general, this is due, in part, to the fact that at high sensation levels normal children obtain scores close to the test ceiling.
7. The difference between the test-retest scores is small on all three tests, particularly at the higher sensation levels.
8. The mean scores of the various lists for each of the three tests used are highly similar, indicating that the several lists of each of the three tests are essentially equivalent.

Recommendations for Further Research

1. An investigation should be conducted comparing articulation scores obtained with the test pictures Ross and Lerman utilized in their standardizing study and also in the present study in contrast to the pictures used in the published form of the test. Such a study would appear essential in interpreting WIPI test data for the commercial form of the test.
2. A study utilizing a 40 dB sensation level for the three tests used in this investigation would provide additional information on the articulation functions of these tests with young children.

3. Studies comparing other multiple-choice picture tests, such as the Discrimination by Identification of Pictures, with the WIPI test would provide clinically valuable information.

BIBLIOGRAPHY

BIBLIOGRAPHY

Barr, B. Pure Tone Audiometry for Pre-School Children. Acta Oto-Laryngologica, Supplement 121 (1955).

Barr, B. Nonorganic Hearing Problems in School-Children. Acta Otolaryngologica, 52, 337-346 (1960).

Barr, B., and Lundstrom, R. Deafness Following Maternal Rubella. Acta Otolaryngologica, 53, 413-423 (1961).

Berk, R., and Feldman, A. Functional Hearing Loss in Children. New England Journal of Medicine, 259, 214-216 (1958).

Brooks, R., and Goetzinger, C. P. Vocabulary Variables and Language Skills in the PB Discrimination of Children. Journal of Auditory Research, 6, 357-370 (1966).

Brockman, S., and Hoversten, G. Pseudo Neural Hypacusis in Children. Laryngoscope, 70, 825-839 (1960).

Carhart, R. Speech Audiometry in Clinical Evaluation. Acta Otolaryngologica, 41, 18-42 (1952).

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

Chaiklin, J., Ventry, I., Barrett, L., and Skalbeck, G. Pure-Tone Threshold Patterns Observed in Functional Hearing Loss. Laryngoscope, 69, 1165-1179 (1959).

Clausen, J. Threshold for Pure Tone and Speech in Retardates. American Journal of Mental Deficiency, 70, 556-562 (1966).

Downs, M., and Hemenway, W. Report on the Screening of 17,000 Neonates. International Audiology, 8, 72-76 (1969).

Dix, M., and Hallpike, C. The Peep Show. British Medical Journal, 2, 719-723 (1947).

Dix, M., and Hallpike, C. Further Observations Upon the Diagnosis of Deafness in Young Children. British Medical Journal, 7 (1947).

Dixon, R., and Newby, H. Children with Nonorganic Hearing Problems. Archives of Otolaryngology, 70 (1959).

Egan, J. Articulation Testing Methods. ORSD Report 3802, United States Department of Commerce, Office of Technical Services, Washington, D.C. (1944).

Fairbanks, G. Test of Phonemic Differentiation: The Rhyme Test. Journal of the Acoustical Society of America, 30, 596-600 (1958).

Fisch, L. Causes of Congenital Deafness. International Audiology, 8, 85-89 (1969).

Fletcher, H., and Steinberg, J. Articulation Testing Methods. Bell System Technical Journal, 8, 806-854 (1929).

Flottorp, G., Morley, D., and Skatvedt. The Location of Hearing Impairment in Athetoids. Acta Otolaryngologica, 48, 404-414 (1957).

Flower, R., Viehweg, W., and Ruzicka, W. The Communicative Disorders of Children with Kernicteric Athetosis: Auditory Disorders. Journal of Speech and Hearing Disorders, 31, 41-59 (1966).

Frisina, D. American Annals of the Deaf, 107, 478-486 (1962).

Froeschels, E. Psychic Deafness in Children. Archives of Neurological Psychiatry, 51, 544-549 (1944).

Geyer, M., and Yankauer, A. Journal of Exceptional Children, 72, 723-735 (1957).

Giolas, T., and Epstein, A. Comparative Intelligibility of Word Lists and Continuous Discourse. Journal of Speech and Hearing Research, 6, 349-358.

Guilford, F., and Haug, C. "Diagnosis of Deafness in the Very Young Child. Archives of Otolaryngology, 55, 101-107 (1952).

Haskins, H. A Phonetically Balanced Test of Speech Discrimination for Children. Unpublished Masters Thesis, Northwestern University, 1949.

Haspiel, G. Picture Identification Test of Discrimination for Speech. Unpublished Ph.D. Dissertation, Pennsylvania State University, 1961.

Haug, O., and Guilford, F. Hearing Testing on the Very Young Child. Trans. Amer. Acad. of Ophth. and Otolaryn., 4, 269-271 (1960).

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

Hirsh, I., Davis, H., Silverman, S., Reynolds, E., Eldert, E., and Benson, R. Development of Materials for Speech Audiometry. Journal of Speech and Hearing Disorders, 17, 321-337 (1952).

Hudgins, C. A Method of Appraising the Speech of the Deaf. Volta Review, 5, 579-601 (1949).

Hutton, C., and Weaver, J. PB Intelligibility and Word Familiarity. Laryngoscope, 69, 1443-1450 (1959).

Keaster, J. A Quantitative Method of Testing the Hearing of Young Children. Journal of Speech and Hearing Disorders, 12, 159-162 (1947).

Kodman, F. Sensory Processes and Mental Deficiency. In Handbook of Mental Deficiency. Edited by N. Ellis. New York: McGraw-Hill (1963).

Kopra, L., Blosser, D., and Waldron, D. Comparison of Fairbanks Rhyme Test and CID Auditory Test W-22 in Normal and Hearing-Impaired Listeners. Journal of Speech and Hearing Research, 11, 735-739 (1968).

Landes, B. Speech Discrimination Testing with Hearing Young Children. Journal of Speech and Hearing Research, 3, 248 (1960).

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

Lerman, J., Ross, M., and McLauchlin, R. A Picture-Identification Test for Hearing-Impaired Children. Journal of Auditory Research, 5, 273-278 (1965).

Lowell, E., Rushford, G., and Hovarsten, R. Evaluation of Pure Tone Audiometry with Preschool Children. Journal of Speech and Hearing Disorders, 21, 292-302 (1956).

Markle, D., and Miller, M. Nature of Deafness in Athetoid Cerebral Palsy. Archives of Otolaryngology, 78, 794-496 (1963).

McNamee, J. An Investigation of the Use of the CID Auditory Test W-22 with Children. Unpublished Masters Thesis, Ohio State University, 1960.

Meyerson, L. A Verbal Audiometric Test for Young Children. American Psychologist, 2, 291-296 (1947).

Miller, M., and Rabinowitz, M. Audiological Problems Associated with Pre-Natal Rubella. International Audiology, 9, 90-98 (1969).

Monsees, E. Children's Auditory Test. The Volta Review, 446 (1953).

Myatt, B., and Landes, B. Assessing Discrimination Loss in Children. Archives of Otolaryngology, 77, 359-362 (1963).

Nasca, F. An Investigation of the PICSI Test: Picture Identification for Children--a Standard Index--Its Standardization Upon Children with Sensori-Neural Hearing Losses and Delayed Language Development. Unpublished Ph.D. Dissertation, Indiana University, 1964.

Nielson, K. Speech Sound Discrimination of Pre-School Children as Measured by the CID Auditory Test W-22. Unpublished Masters Thesis, Michigan State University, 1960.

O'Neill, J., Oyer, H., and Hillis, J. Audiometric Procedures Used with Children. Journal of Speech and Hearing Disorders, 26, 61-66 (1961).

Owens, E. Intelligibility of Words Varying in Familiarity. Journal of Speech and Hearing Research, 4, 113-129 (1961).

Oyer, H., and Doudna, M. Structural Analysis of Word Responses Made by Hard of Hearing Subjects on a Discrimination Test. Archives of Otolaryngology, 70, 357-364 (1959).

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

Peterson, J. Nonorganic Hearing Loss in Children and Bekesy Audiometry. Journal of Speech and Hearing Disorders, 28, 153-158 (1963).

Rintelmann, W., and Harford, E. The Detection and Assessment of Pseudohypoacousis Among School-Age Children. Journal of Speech and Hearing Disorders, 28, 141-152 (1963).

Rintelmann, W., and Jetty, A. Reliability of Speech Discrimination Testing Using CNC Monosyllabic Words. Unpublished study, Michigan State University, 1968.

Rintelmann, W. and Burchfield, S. Reliability of Speech Discrimination Testing Using CNC Monosyllabic Words: A Follow-up Study. Unpublished study, Michigan State University, 1970.

Rosenzweig, M., and Portman, L. Intelligibility as a Function of Frequency of Usage. Journal of Experimental Psychology, 54, 412-422 (1957).

Ross, M., Huntington, D., Newby, H., and Dixon, R. Speech Discrimination of Hearing-Impaired Individuals inNoise: It's Relationship to Other Audiometric Parameters. Journal of Auditory Research, 5, 47-72 (1965).

Ross, M., and Lerman, J. A Picture Identification Test for Hearing-Impaired Children. Journal of Speech and Hearing Research, 13, 44-53 (1970).

Seidel, S. The PICSI Test: Picture Identification for Children--a Standardized Index. Unpublished Masters Thesis, Indiana University, 1963.

Siegenthaler, B., and Haspiel, G. Development of Two Standardized Measures of Hearing for Speech by Children. U.S. Department of Health, Education and Welfare, Project No. 2372, Contract No. OE-5-10-003, 1966.

Siegenthaler, B., Pearson, J., and Lezak, R. A Speech Reception Threshold Test for Children. Journal of Speech and Hearing Disorders, 19, 360-366 (1954).

Sims, A. Word List Picture Identification Test for Speech-Hearing in Preschool Children. Unpublished Masters Thesis, Louisiana State University, 1961.

Sortini, A., and Flake, C. Speech Audiometry for Pre-School Children. Laryngoscope, 63, 991-997 (1953).

Tillman, T., Carhart, R., and Wilber, L. A Test for Speech Discrimination Composed of CNC Monosyllabic Words: N.U. Auditory Test #4, Technical Documentary Report No. SAM-TDR-62-135, 1963.

Tillman, T., and Carhart, R. An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words: Northwestern University Auditory Test #6, Technical Documentary Report No. SAM-TR-66-55, 1966.

Tillman, T., Johnson, R., and Olsen, W. Earphone Versus Sound-Field Threshold Sound-Pressure Levels for Spondee Words. Journal of the Acoustical Society of America, 39, 125-133 (1966).

Utley, J. What's Its Name. Urbana, Illinois: University of Illinois Press, 1951, pp. 126-127.

APPENDIX

Individual data for 3½ year old children during the test session, Age 3½ - Test.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent WIPI 8	16	24	32	SRT	Discrimination Score in Percent PBK-50 16	24	32
1	F	42	12	--	80	96	96	10	--	--	78
2	F	43	14	--	84	96	88	12	--	--	90
3	M	42	14	--	76	88	84	14	--	--	78
4	F	44	6	--	76	84	88	4	--	--	82
5	F	41	16	--	92	76	96	18	--	--	60
6	M	44	8	--	96	96	100	12	--	--	84
7	F	40	12	--	64	96	88	8	--	--	54
8	M	43	12	--	84	96	96	14	--	--	64
9	M	42	14	--	68	88	84	14	--	--	44
10	F	39	18	--	64	92	96	14	--	--	72
11	M	44	14	--	84	92	100	16	--	--	96
12	F	41	14	--	76	60	84	16	--	--	58

Individual data for 3½ year old children during the retest session. Age 3½ - Retest.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent WIPI 8	16	24	32	SRT	Discrimination Score in Percent PBK-50 16	24	32
1	F	42	--	--	--	96	--	--	--	--	70
2	F	43	--	--	--	88	--	--	--	--	86
3	M	42	--	--	--	--	--	--	--	--	64
4	F	44	--	--	--	80	--	--	--	--	82
5	F	41	--	--	--	80	--	--	--	--	50
6	M	44	--	--	--	100	--	--	--	--	74
7	F	40	--	--	--	88	--	--	--	--	60
8	M	43	--	--	--	88	--	--	--	--	76
9	M	42	--	--	--	96	--	--	--	--	46
10	F	39	--	--	--	92	--	--	--	--	78
11	M	44	--	--	--	100	--	--	--	--	92
12	F	41	--	--	--	56	--	--	--	--	56

Individual data for 5½ year old children during the test session. Age 5½ - Test.

Subject	Sex	Age in Months	Discrimination Score in Percent WIPI SRT	8	16	24	32	Discrimination Score in Percent PBK-50 SRT	16	24	32
1	M	65	8	76	96	100	92	6	80	94	94
2	M	67	2	88	80	96	96	2	88	90	96
3	F	68	8	84	100	92	100	4	90	96	96
4	F	64	6	88	100	96	100	4	86	98	98
5	F	68	10	80	100	92	96	8	94	96	100
6	F	64	10	84	100	100	100	8	90	96	96
7	M	64	8	92	96	96	96	6	84	88	96
8	M	65	10	88	100	96	96	12	88	92	90
9	F	66	6	96	88	96	100	4	86	92	96
10	M	64	10	68	88	92	96	14	86	88	96
11	F	68	10	80	80	92	96	10	82	88	94
12	M	68	8	72	100	100	100	8	76	96	98

Individual data for 5½ year old children during the retest session. Age 5½ - Retest.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent (WIPI) 8	16	24	32	SRT	Discrimination Score in Percent (PBK-50) 16	24	32
1	M	65	--	92	96	100	96	--	88	94	96
2	M	67	--	80	96	100	92	--	90	92	96
3	F	68	--	96	92	96	100	--	88	98	96
4	F	64	--	92	100	100	100	--	82	94	100
5	F	68	--	92	100	92	100	--	92	96	98
6	F	64	--	88	96	100	100	--	84	98	98
7	M	64	--	96	100	96	100	--	84	94	98
8	M	65	--	80	100	96	92	--	88	90	94
9	F	66	--	92	96	100	100	--	96	94	98
10	M	65	--	96	84	88	100	--	80	90	88
11	F	68	--	56	88	76	100	--	80	86	90
12	M	68	--	92	100	100	100	--	84	84	96

Individual data for 7½ year old children during the test session. Age 7½ - Test.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent WIPI 8	16	24	32	SRT	Discrimination Score in Percent PBK-50 16	24	32	SRT	Discrimination Score in Percent N.U. Auditory Test No. 6 8	16	24	32
1	M	87	2	84	88	96	92	4	92	96	98	0	40	74	82	90
2	M	90	8	92	100	96	100	4	82	96	96	6	62	78	88	88
3	F	88	12	100	96	100	100	12	90	94	98	14	56	84	96	98
4	F	88	8	92	100	96	100	6	96	96	98	2	72	88	90	98
5	M	92	4	80	100	96	100	0	94	100	98	- 2	44	86	84	98
6	F	88	6	84	92	96	100	4	86	94	98	6	60	86	92	92
7	M	91	2	96	92	96	96	6	92	96	96	4	66	76	96	94
8	F	90	0	72	92	100	100	2	88	94	98	2	46	82	90	88
9	F	90	- 2	100	92	92	100	4	92	94	98	- 2	36	76	84	92
10	F	90	8	96	100	100	100	0	94	96	100	4	54	84	90	92
11	F	92	0	92	100	1-0	96	- 4	82	94	98	- 2	60	86	94	94
12	M	89	6	84	92	100	100	4	86	90	94	4	80	76	80	90

Individual data for 7½ year old children during the retest session. Age 7½ - Retest.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent WIPI 8	16	24	32	SRT	Discrimination Score in Percent PBK-50 16	24	32	SRT	Discrimination Score in Percent N.U. Auditory Test No. 6 8	16	24	32
1	M	87	--	92	96	100	92	--	94	96	100	--	52	76	76	90
2	M	90	--	96	96	100	96	--	84	94	96	--	60	76	90	92
3	F	88	--	100	96	100	100	--	92	100	96	--	38	88	98	92
4	F	88	--	96	100	100	100	--	88	98	98	--	60	82	88	90
5	M	92	--	100	100	96	100	--	90	94	100	--	36	68	90	98
6	F	88	--	92	96	100	100	--	92	86	98	--	54	82	90	94
7	M	91	--	92	96	100	100	--	92	92	98	--	54	88	96	96
8	F	90	--	76	100	92	100	--	92	96	98	--	42	70	84	90
9	F	90	--	80	96	96	100	--	90	90	96	--	50	76	90	90
10	F	90	--	92	96	100	100	--	90	96	100	--	44	84	82	94
11	F	92	--	92	100	100	100	--	80	88	96	--	54	76	92	98
12	M	89	--	76	96	100	100	--	88	88	96	--	78	72	82	86

Individual data for 9½ year old children during the test session. Age 9½ - Test.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent WIPI 8	16	24	32	SRT	Discrimination Score in Percent PBK-50 16	24	32	SRT	Discrimination Score in Percent N.U. Auditory Test No. 6 8	16	24	32
1	M	114	2	72	96	100	100	2	82	94	98	2	60	70	84	90
2	M	112	4	88	96	96	100	6	90	94	100	8	64	88	92	96
3	F	111	4	96	96	100	100	2	88	96	98	6	76	82	92	98
4	M	116	6	92	96	100	100	6	86	98	98	6	58	78	92	94
5	M	112	6	92	96	100	100	6	92	100	100	2	62	88	94	98
6	F	111	4	88	100	100	100	4	94	96	100	0	56	74	88	92
7	F	111	2	84	100	96	100	4	94	92	96	0	64	82	88	92
8	M	113	8	96	100	96	96	8	94	98	100	8	68	88	86	92
9	M	115	- 2	100	100	96	100	- 2	92	98	98	- 2	54	74	80	94
10	M	115	4	88	96	100	100	4	84	98	98	2	62	82	92	88
11	F	113	2	80	88	100	96	6	88	90	98	6	62	72	86	94
12	M	115	12	96	96	100	96	14	86	96	98	12	58	80	98	94

Individual data for 9½ year old children during the retest session. Age 9½ - Retest.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent WIPI 8	16	24	32	SRT	Discrimination Score in Percent PBK-50 16	24	32	SRT	Discrimination Score in Percent N.U. Auditory Test No. 6 8	16	24	32
1	M	114	--	88	96	100	100	--	76	90	100	--	60	80	84	90
2	M	112	--	92	96	96	100	--	94	98	98	--	70	88	94	100
3	F	111	--	100	96	100	100	--	84	98	100	--	78	92	98	100
4	M	116	--	96	100	100	96	--	90	100	98	--	74	90	94	96
5	M	112	--	92	96	100	100	--	96	100	100	--	50	64	94	100
6	F	111	--	92	96	100	100	--	98	98	98	--	60	84	96	98
7	F	111	--	92	100	96	100	--	90	98	96	--	66	72	94	92
8	M	113	--	92	100	100	100	--	92	98	100	--	72	80	86	90
9	M	115	--	88	100	100	96	--	92	94	96	--	54	82	84	96
10	M	115	--	84	92	100	100	--	88	94	98	--	58	78	94	92
11	F	113	--	80	96	96	100	--	84	92	96	--	44	82	90	94
12	M	115	--	88	100	100	100	--	86	96	100	--	52	76	90	96

Individual data for 11½ year old children during the test session. Age 11½ - Test.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent PBK-50 16	24	32	SRT	Discrimination Score in Percent N.U. Auditory Test No. 6 8	16	24	32
1	F	138	10	72	98	100	12	50	74	82	100
2	F	140	0	82	100	98	2	68	88	94	98
3	F	139	12	96	100	100	8	62	84	94	98
4	F	137	2	90	98	100	0	66	72	78	92
5	M	142	0	92	98	98	0	56	78	92	100
6	F	142	- 2	94	96	98	- 2	64	78	92	94
7	F	139	4	98	98	100	6	72	84	90	96
8	F	141	4	90	100	100	2	68	86	92	92
9	F	139	10	90	94	98	14	54	86	90	96
10	F	138	4	92	92	100	4	58	86	94	98
11	M	137	2	94	98	100	2	70	94	94	98
12	F	138	- 4	80	92	94	- 2	64	78	74	96

Individual data for 11½ year old children during the retest session. Age 11½ - Retest.

Subject	Sex	Age in Months	SRT	Discrimination Score in Percent PBK-50 16	24	32	SRT	Discrimination Score in Percent N.U. Auditory Test No. 6 8	16	24	32
1	F	138	--	78	98	100	--	58	80	88	100
2	F	140	--	80	100	100	--	60	76	92	98
3	F	139	--	96	100	100	--	66	80	98	98
4	F	137	--	96	98	98	--	56	78	92	94
5	M	142	--	92	100	100	--	72	86	90	98
6	F	142	--	96	98	98	--	64	82	96	94
7	F	139	--	94	100	100	--	74	86	94	94
8	F	141	--	86	100	100	--	76	86	94	96
9	F	139	--	88	98	100	--	76	94	88	94
10	F	138	--	90	90	100	--	66	92	96	96
11	M	137	--	98	96	100	--	82	94	98	96
12	F	138	--	82	90	96	--	62	76	82	94