RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 21 (1996-1997)
Indiana University

Performance of Normal-Hearing Children on Open-Set Speech Perception Tests[1]

Melissa Kluck,[2] David B. Pisoni,[3] and Karen Iler Kirk[3]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This research was supported by NIH Research Grant DC-00111 to Indiana University in Bloomington, IN.
[2] Now at the Department of Communication Sciences and Disorders, Northwestern University, Chicago, IL
[3] Also DeVault Otologic Laboratory, Indiana University School of Medicine, Indianapolis, IN.

Performance of Normal-Hearing Children on Open-Set Speech Perception Tests

Abstract. The Phonetically Balanced Kindergarten Test (PBK), an open-set test of word recognition is typically included in test batteries designed to assess the speech perception skills of profoundly deaf children with cochlear implants. Many pediatric cochlear implant users have a great deal of difficulty with this test. Two new open set tests, the Lexical Neighborhood Test (LNT), and Multisyllabic Lexical Neighborhood Test (MLNT) (Kirk, Pisoni & Osberger, 1995), have been developed with the framework of the Neighborhood Activation Model (NAM) (Luce, 1986) of spoken word recognition. The LNT and MLNT are based on the lexical characteristics of word frequency and neighborhood density, and include words found in the vocabularies of children age three to five. Results from these tests with pediatric cochlear implant users have shown that their lexicons appear to be organized into similarity neighborhoods, and these neighborhoods are accessed in open-set word recognition tests. The present study investigates the speech perception abilities of normal hearing children ages three and four using the PBK, LNT and MLNT. The study was also designed to assess test-retest reliability for these tests using normal hearing children. Each child was first screened to ensure normal hearing, and measures of each child's vocabularies were obtained. Normal hearing three and four year old children performed extremely well, with scores near ceiling on all three tests. Because of the lack of variance in the scores, test-retest reliability could not be assessed using this population. The study did demonstrate, however, that it is reasonable to expect normal hearing three- and four-year old children to recognize all the words from these three open-set speech perception tests at very high levels of performance. These results can be used as a benchmark for children with hearing impairments who score poorly on the PBK, LNT and MLNT.

Introduction

This study is concerned with assessing the spoken word recognition abilities of children three to four years old. Several justifications exist for testing speech perception abilities of young children. First, these measures can aid in assessing rate of language development in normal as well as hearing or language impaired children, and are often used to detect and diagnose language delay. Second, a child's speech perception abilities may also provide insight about perceptual and cognitive skills important for language and reading. In the case of hearing impaired children, speech perception tests are critical in assessing progress in language skills, reading, and overall cognitive development, and as aids in planning education, rehabilitation, and speech therapy (Kirk, Diefendorf, Pisoni & Robbins, 1997).

Over the years, two types of word recognition tests have been used with children: closed-set tests and open-set tests. Closed-set tests require a child to choose between several possible responses, whereas in open-set tests the potential responses are unlimited. Closed-set tests are useful when testing perceptual abilities of children who cannot speak clearly or write, because the procedures force the subject to respond by pointing to pictures. Open-set tests are not appropriate for all children. Successful performance on open-set tests requires that the child be able to hear well enough to perceive the stimulus pattern, encode it, and

then represent it in memory, and speak intelligibly enough so that the examiner can understand the responses when the child attempts to reproduce it at the time of test (Kirk et al., 1997).

Since the late 1940's, the Phonetically Balanced Kindergarten test (PBK) (Haskins, 1949), an open-set, monosyllabic speech perception test, has been used for testing young children who are profoundly deaf. More recently, this test has been used with children who have received cochlear implants (Kirk et al., 1997). The test consists of three lists, each with fifty words each, chosen from the International Kindergarten Union vocabulary lists. Each list is phonetically balanced. That is, the frequency of different phonemes reflects phoneme distribution in normal conversational speech.

With the exception of a small number of pediatric cochlear users who perform exceptionally well on this test, the so-called "Stars", the PBK is extremely difficult for many deaf children (Pisoni, Svirsky, Kirk & Miyamoto, 1997). Despite the difficulty of this test, it is routinely included in the standard battery of tests used to assess progress of cochlear implant users at Central Institute for the Deaf (CID), Indiana University School of Medicine, and many other cochlear implant centers around the world.

Although the PBK was originally designed for normal-hearing young children, and has been used by clinical audiologists for many years, it was not based on any theory or model of speech perception. Instead, it was developed primarily on empirical criteria. Low scores on this test by hearing-impaired children may be due to many factors. For example, the PBK test may be inappropriate for cochlear implant users because the vocabulary is too difficult (Osberger, Miyamoto, Zimmerman-Phillips, Kemick, Stroer, Firszt & Novak, 1991). Alternatively, perhaps deaf children have trouble perceiving isolated words in open-set tests where there are no response alternatives provided. To address these criticisms, specifically the issue about unfamiliar words, Kirk, Pisoni and Osberger (1995) recently developed several new open-set tests of speech perception to assess children using cochlear implants. These tests are firmly grounded in recent theoretical work on spoken word recognition (Luce, Pisoni & Goldinger, 1990). All the words were selected to be highly familiar to young children. The Lexical Neighborhood Test (LNT) contains two lists of 50 mono-syllabic words. The Multisyllabic Lexical Neighborhood Test (MLNT) contains two lists of 24 words, which are 2-3 syllables in length.

These new word recognition tests are theoretically motivated and make a number of specific predictions. That is, they were designed to test certain principles of speech perception and spoken word recognition processes used by normal hearing subjects. The tests were developed to provide descriptive data and theoretical insights into the underlying perceptual processes employed by pediatric cochlear implant users. The tests were also designed to be easier for cochlear implant users to carry out. Because the tests are easier, they should be more useful in discriminating between open-set test abilities of different populations of hearing-impaired children.

The LNT and MLNT tests are based on the Neighborhood Activation Model (NAM) (Luce, 1986) of spoken word recognition. The Neighborhood Activation Model assumes that two important characteristics of a word are its frequency (i.e., how often a word appears in the language), and its lexical neighborhood, (i.e., the number of words which are phonetically similar to it). These two factors have been shown to affect spoken word recognition when isolated words are presented to adults in noise (Luce, 1986; Goldinger, Luce, & Pisoni, 1989; Luce et al., 1990). A "lexical neighborhood" refers to the number of words which differ from a target word by only one phoneme. For example, the lexical neighbors of "bat" include the words *"cat", " ban"*, and *"bath"*. If a word has many lexical neighbors, it is assumed to reside in a "dense" lexical neighborhood. In contrast, if a word has few lexical neighbors, it is assumed to reside in a "sparse" lexical neighborhood.

Luce (1986) and Luce et al. (1990) have shown that normal hearing adult listeners find it easier to recognize high frequency words from sparse lexical neighborhoods (i.e., "easy" words) than low frequency words from dense lexical neighborhoods (i.e., "hard" words). The Neighborhood Activation Model has received a great deal of support from word recognition studies using adult listeners (Luce et al., 1990). The most important theoretical claim of the NAM model is that spoken words are recognized relationally in the context of other words in the listener's mental lexicon using a two-step process of acoustic-phonetic activation followed by lexical selection. Frequency and acoustic-phonetic similarity play independent roles in this overall process.

Lexical characteristics of words also have been shown to affect word recognition in adults with mild to moderate hearing loss (Pisoni, Miyamoto, Kirk, Sommers & Osberger, 1994). These findings encouraged the development of a test based on lexical characteristics of words to be used for testing children with profound hearing loss who are using cochlear implants. The LNT and MLNT tests were designed to measure speech perception skills and to provide new information about the way in which children with cochlear implants organize and access spoken words from memory. Differences in lexical neighborhood density and word frequency were used to generate lists of "easy" and "hard" words for these new tests, based on lexical characteristics of each word using assumptions derived from the NAM model (Kirk et al,. 1995).

The LNT contains two "easy" and two "hard" word lists. All words in this test are monosyllabic and exclude proper nouns, possessives, contractions, plurals and inflected forms. Words were chosen according to mean ratings of frequency and neighborhood density from Logan's (1992) earlier analysis of a large database of children's utterances. Logan used the CHILDES (Child Language Data Exchange System) database (MacWhinney & Snow, 1985), and computed lexical characteristics of words contained in the vocabularies of children aged one to five. Test items used for the LNT and MLNT were selected from words produced by children (ages three to five) contained in Logan's (1992) analyses. These ratings were calculated using the utterances analyzed by Logan. "Easy" words had frequencies above the median, and neighborhood densities below the median. The opposite is true for "hard" words. The MLNT contains two "easy" and two "hard" words, each with 12 words. These stimuli were chosen in a manner identical to that of all words included in the LNT, except the words are multi-syllabic.

Kirk et al. (1995) reported significant differences in lexical characteristics of the PBK, LNT and MLNT word lists, when rated for familiarity by normal hearing adult listeners. Although all three word lists were found to be highly familiar to adult listeners, important differences were found in neighborhood density. LNT "hard" words had the highest scores for neighborhood density, followed PBK words, and then LNT "easy" words. Using a computerized database, Kirk et al. (1995) investigated how familiar the PBK words were to young children, and found that only 31% of the PBK words were included in he CHILDES database analyzed by Logan (1992), which suggests that many of the PBK words are not common in the vocabularies of children under the age of five. This result suggests that the vocabulary used in the PBK may be unfamiliar to young children and pediatric cochlear implant users, which may be one of the reasons many deaf children with cochlear implants perform so poorly on this test. All of the LNT and MLNT words were chosen from words known to be in the vocabularies of children aged 1 to 5. Thus, these familiar words are more likely to be in the vocabularies of young children.

When pediatric cochlear implant users were assessed using the PBK, LNT and MLNT, Kirk et al. (1995) found significant differences in correct identification of the words on the three word lists. Word length and lexical characteristics had a significant effect of identification scores. Specifically, subjects'

performance was highest on the MLNT, followed by LNT, and then PBK. Within the LNT and MLNT, scores on the "easy" lists were consistently higher than scores on the "hard" lists.

Kirk et al. (1995) concluded that the lexical characteristics of words on these tests affects word recognition scores of pediatric cochlear implant users. Based on the observed pattern of scores across the LNT and MLNT tests, Kirk et al. (1995) suggested that the lexicons of pediatric cochlear implant users are organized in a manner that is similar to normal hearing children and adults. That is, words are organized in long term memory according to similarity neighborhoods and frequency. Kirk et al. (1995) also found that longer words i.e. MLNT were easier to identify than short words (i.e., LNT), suggesting that these hearing-impaired children use the length of a word to discriminate and select spoken words from lexical memory.

Because no baseline data have been obtained yet for the LNT and MLNT using normal hearing children, we administered the PBK, LNT and MLNT to normal hearing three- and four-year old children. We also collected data on their vocabulary knowledge. The present study had two primary goals. The first goal was to examine test-retest reliability of the LNT and MLNT word lists. Test-retest reliability can be established by repeating the administration of an identical test after a reasonable amount of time has passed (Anastasi, 1961) to determine if the scores are repeatable. The second goal of the study was to provide normal hearing control data for the LNT and MLNT test. We expected to find effects of age on performance for these word recognition tests, such that older children should do better than younger subjects. We were also interested in knowing at what age is it reasonable to use the LNT and MLNT with children who have some hearing loss. We addressed these problems using two groups of young children who had normal hearing and no known language or other developmental delays.

Methods

Participants

All subjects were recruited from a database of children in the Bloomington, Indiana area. Letters and subsequent follow-up phone calls were used to inform parents of the experiment. Two groups of subjects were used. Eight four-year olds, 2 males and 6 females with an average age of 4 years 9 months participated in one group. One four-year old male did not return for the second session. Twenty-two three-year olds also participated in the other group: 8 males and 14 females, with an average age of 3 years 4 months. One three- year old female did not return for the second session. One three-year old male did not complete the first session, and one three-year old female did not complete either session. Only scores from completed tests were used in the final data analysis.

Test Materials

Three sets of words were used in constructing the audio recordings of the test stimuli. The first set was made up of the three PBK 50-word lists. The second set of words was made up of both LNT 50-word lists, each containing 25 "easy" and 25 "hard" words. The third set was made up of both MLNT 24-word lists, each containing 12 "easy" and 12 "hard" words. Thus, the total number of test stimuli was 298: 150 PBK, 100 LNT and 48 MLNT words.

The test words were read in isolation by a male talker. The stimuli were recorded in a single walled sound attenuated chamber (IAC No. 402) using a head mounted close talking microphone (Shure SM98A). The analog signal was low-pass filtered at 10 kHz (TDT FT5) and digitized directly to disk at 22 kHz using 16 bit resolution with the TDT System II interfaced to a 486 DX-66 based PC. Following the digital

recordings, the waveform files were down-sampled to 20 kHz. To ensure uniform presentation levels, overall RMS (root mean squared) amplitudes were digitally equated for all the stimulus files.

Procedure

Subjects were screened on the first day of the study for normal hearing at 250, 500, 1,000, 2,000 and 4,000 Hz at 20 dB HL using a Maico Hearing Instruments audiometer (MA27) and TDH-39P headphones. Both ears were screened separately. Vocabulary scores were then collected using the revised version of the Peabody Picture Vocabulary Test (PPVT-R) (Dunn & Dunn, 1981).

The perceptual data for the three word recognition tests were collected in two separate sessions that were conducted between 2 and 9 days apart, with a mean time lapse of 5.0 days (SD = 2.152) for the three-year olds, and between 2 and 6 days apart with mean time lapse of 3.7 days (SD = 1.7) for the four-year olds. In each session, subjects heard one of three 50-word PBK lists, counterbalanced across subjects. Then, the LNT and MLNT lists were presented, in a counterbalanced order across subjects. Within the LNT and MLNT word sets, presentation order of lists was also counterbalanced. Stimuli were randomized within all lists. Subjects heard the same lists in the same order in the second session.

An IBM Thinkpad Computer (Model 750CS) was used to playback stimuli and record responses entered by the examiner. Stimuli were reproduced using a loudspeaker (Acoustics Research Instruments). The output level of the speech was set at 70 dB SPL, as determined by a hand-held sound pressure level meter (Triplett 370) that was placed at the approximate position of the subjects' head. Responses were transcribed immediately by the experimenter, using a computer program specifically designed to present stimuli, and record the examiner's transcription of the subject's response. A tape recorder was also used to record all subjects' responses during a test session.

During the first session, parents completed the Language Development Survey (Rescorla, 1991), and the MacArthur Communicative Development Inventory (MacArthur, 1993). These scales are used to detect language delay. During the second session, parents completed a familiarity task (FAM) of their child's vocabulary using the PBK, LNT, and MLNT word lists. This task required the parent to rate their child's familiarity with each of the test words on a scale from 1 to 7. Each parent was also asked to complete a survey of their own vocabulary knowledge (PFAM). Four hundred and fifty words were chosen from the Hoosier Mental Lexicon (Nusbaum, Pisoni & Davis, 1984), which contains familiarity ratings for 20,000 words chosen from the *Merriam-Webster Pocket Dictionary*. Based on the Hoosier Mental Lexicon, 150 words were rated as highly familiar, 150 had medium familiarity ratings, and 150 had low familiarity ratings. Parents were asked to rate their familiarity with these words on a scale from 1 to 7.

Children received stickers and candy to motivate them to continue during the experiment. At the end of the first session, each child received a small toy. Upon completion of the experiment, subjects were paid $10, and received a T-shirt with the laboratory logo, and a certificate of participation.

Results

First, we will report demographic data for the subjects. Next we will consider the mean scores for the subjects on the three word recognition tests. Following this we discuss the test-retest reliability pertaining to the word recognition scores. Finally, we will look at the familiarity measures completed by the parents. In each section, data are separated by age group. Scores for the three-year old group are followed by scores from the four-year old group for all measures reported below.

Demographic Data

Table 1 displays demographic data for both groups of subjects. Within the three-year old group, twenty-two subjects passed the hearing screening on day 1, and participated in the first session. The average age for this group was 40 months (SD = 3.7). Scores from the Peabody Picture Vocabulary Test (PPVT-R) had a mean of 44.5 (SD = 10.0). Within the three-year old group, chronological age and PPVT-R scores were correlated ($r = +0.63$, $p < .005$).

Within the four-year old group, eight subjects passed the hearing screening and participated in the first session. The average age of this groups was 57 months (SD = 1.2). The average of the PPVT-R scores, reported in terms of mental age, was 62.6 (SD = 13.3).

Table 1

Demographic Data for Both Age Groups

	Mean (mo.)	SD (mo.)	N	PPVT-R (mo.)	SD (mo.)
3 year olds	40	3.7	22	44.5	10.0
4 year olds	57	1.2	8	62.6	13.3

Word Recognition Tests

Table 2 displays the means for the perception tests for the three-year old group on both days. Means reported for the LNT and MLNT tests were averaged across lists 1 and 2, so for each test, scores are reported for "easy" and "hard" lists. On both days subjects performed highest on MLNT "easy" words, followed by MLNT "hard" words. The means for these tests were all above 90.5%. Means on the LNT "easy" and LNT "hard" tests ranged between 87.7% for LNT "hard" Day 2, and 89.2% for LNT "easy" Day 1. Means were lowest for the PBK test. The mean for Day 1 was 84.0%, and the mean for Day 2 was 86.8%.

Table 2

Results of Perception Tests for 3 Year Olds

	LNT "easy"	LNT "hard"	MLNT "easy"	MLNT "hard"	PBK
Mean % day 1	89.2% [b]	89.1% [b]	90.7% [a]	90.5% [a]	84.0%
(SD)	8.0	9.1	7.9	10.7	7.9
Mean % day 2	88.4% [b]	87.7% [b]	92.7% [b]	91.5% [b]	86.8% [a]
(SD)	9.3	6.6	8.5	7.6	11.3

Note. [a] N = 21 for these tests. [b] N = 20 for these tests.

Table 3 reports the means for the perception tests on Day 1 and Day 2 for the four-year old group. Means reported for the LNT and MLNT were averaged across lists 1 and 2, as they were for the three-year old group. The means ranged between 96.0% and 99.5%. For both days means were lowest on LNT "hard" words. The mean for Day 1 on this test was 95.7%, and the mean for Day 2 was 96.6 %. The second lowest mean was for the PBK test, with a mean of 96.0% for Day 1, and 96.9% on Day 2. On Day 1 the highest scores were on MLNT "easy" words, with a mean of 99.5%. On Day 2 the highest scores were on MLNT "hard" words, with a mean of 98.8%.

Table 3

Results of Perception Tests for 4 Year Olds

	LNT "easy"	LNT "hard"	MLNT "easy"	MLNT "hard"	PBK
Mean % day 1	97.5%	95.7%	99.5%	98.4%	96.0%
(SD)	2.3	1.7	1.5	3.1	4.8
Mean % day 2 [a]	98.3%	96.6%	97.6%	98.8%	96.9%
(SD)	1.8	3.2	3.3	2.0	2.0

Note. [a] N = 7 for these tests.

Figure 1 displays the means on the PBK, LNT and MLNT tests for both age groups. In this figure, LNT and MLNT scores were averaged across "easy" and "hard" lists, and across lists 1 and 2. The top panel shows the means from Day 1, with the three-year olds on the left and the four-year olds on the right. The bottom panel shows the comparison of means for Day 2. Overall, subjects had the lowest scores on the PBK, followed by the LNT, and then the MLNT. This was true for both the three- and four-year olds on both days. Figure 1 also shows that the four -year olds performed consistently better than the three- year olds, and that there was less variability in the scores from the four- year olds than in the scores from the three- year olds. This figure also shows that scores are close to the ceiling for both groups of subjects.

Insert Figure 1

Figure 2 shows the comparisons between the "easy" and "hard" lists for the LNT and MLNT. Data from the three- and four-year olds are shown separately. The top panel shows the means for Day 1, and the bottom panel shows the means for Day 2. In all but one case, the scores were slightly higher on the "easy" word lists than the "hard" word lists, but in all cases they were consistently close to ceiling levels of performance for both groups of subjects. The one exception was the MLNT test for the four- year old group on Day 2, where the mean was slightly higher on the "hard" words than the "easy" words.

Insert Figure 2

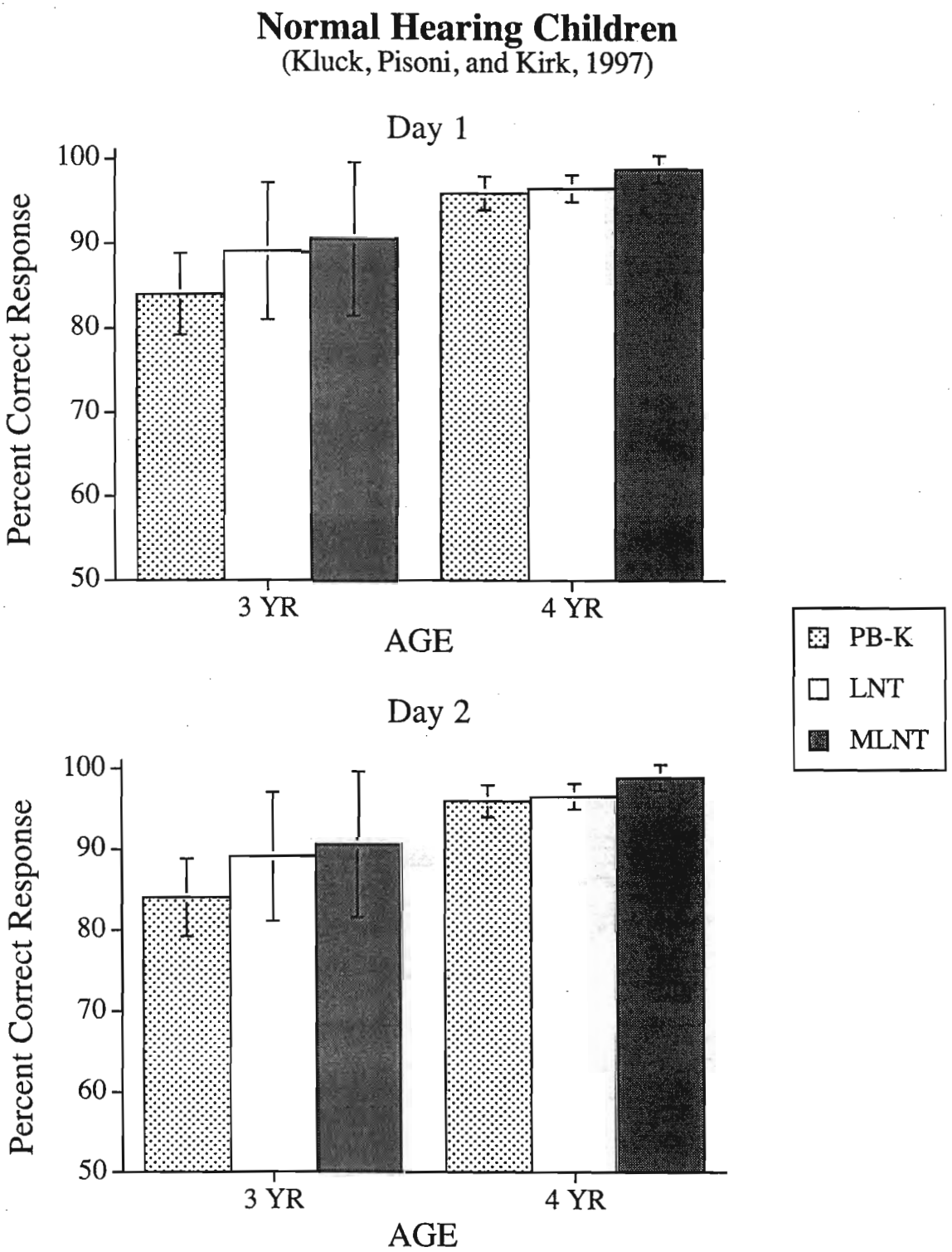

Figure 1. Percentage of words correctly identified for the LNT, MLNT and PBK, with scores broken down by age group, and day of testing.

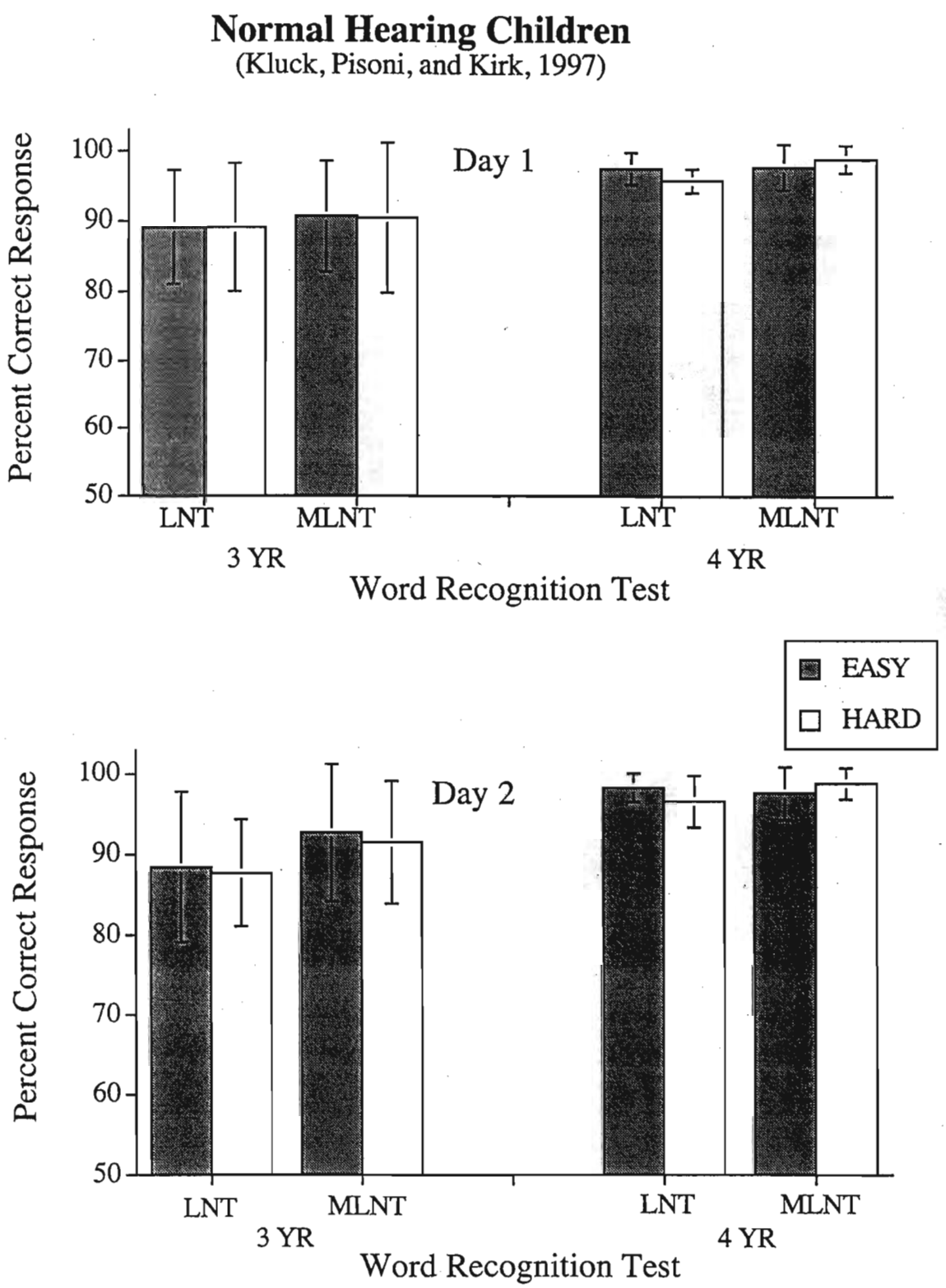

Figure 2. Comparisons between easy and hard word lists for the LNT and MLNT tests, broken down by age group and day of testing.

Paired t-tests were used to find differences between scores on the "easy" and "hard" word lists for each test. The "easy" and "hard" lists were significantly different for the LNT test on day 1 for the 4 year old group. There were no other significant differences between "easy" and "hard" word lists.

Test-retest Correlations

During this study, all subjects participated in two sessions in order to obtain data to assess test-retest reliability for all three perception tests. Because the scores were near ceiling on all of the perception tests, there was not sufficient variation or variability to report meaningful correlations. Figure 3 is a scatterplot of the PBK scores from both age groups, with Day 1 scores plotted using the x-axis, and Day 2 plotted using the y-axis. Figure 4 is a similar scatterplot using the scores from the LNT, and Figure 5 uses the scores from the MLNT. The purpose of these scatterplots is to demonstrate the high levels of performance, as well as the lack of variability in scores on these three tests.

Insert Figure 3

Insert Figure 4

Insert Figure 5

Parent Measures

The parents of the subjects completed several different measures during the course of the study. The Language Development Survey and MacArthur Communicative Development Inventory were used to detect language delay. All subjects knew almost every word on the vocabulary checklists of these measures, and showed adequate sentence formation ability. Therefore, there was no evidence of language delay in any of the participants.

Parents also completed familiarity ratings (FAM) of their child's knowledge of words used in the perception tests. For each word, parents used a rating scale from 1 to 7, with a high score indicating high familiarity with the word, and a low score indicating little or no knowledge of the word. The mean scores from both age groups on the three perception tests are reported in Table 4. The lowest scores in both age group were on the PBK FAM test, indicating that these words are judged by parents to be less familiar to the children than the LNT and MLNT words. The highest ratings for both age groups was on the LNT FAM test. Familiarity ratings for the four- year old group were consistently higher than the ratings for the three- year old group, although a two factor ANOVA showed that age did not have a significant effect on these familiarity ratings. This ANOVA did show an effect of word list ($F = 18.686$, $p<.0001$). Paired t-tests showed significant differences between all three word lists.

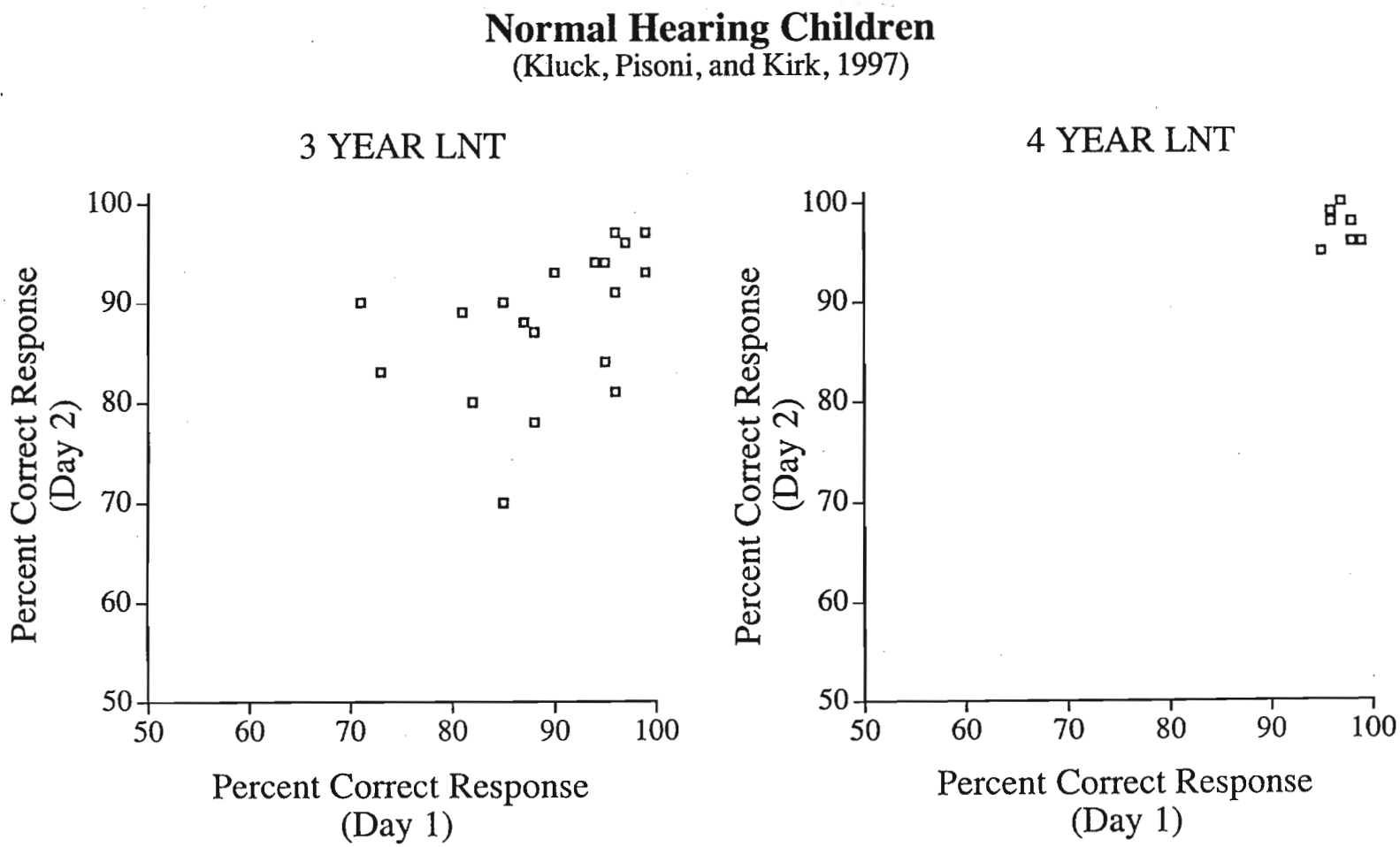

Figure 3. Scatterplots of percent correct scores for PBK for Day 1 and Day 2, with scores for the three-year olds on the left, and scores for the four-year olds on the right.

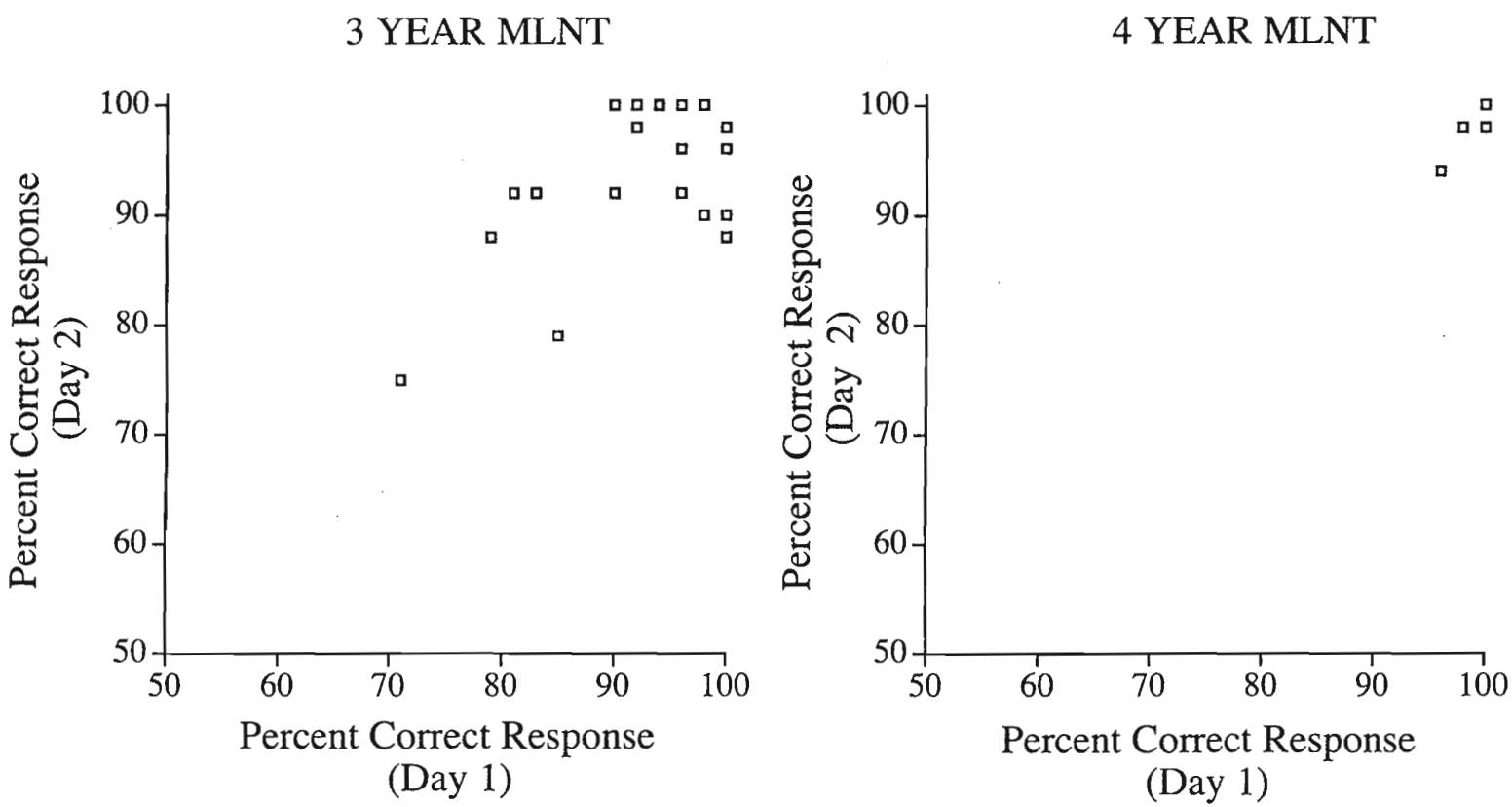

Figure 4. Scatterplots of percent correct scores for LNT for Day 1 and Day 2, with scores for the three-year olds on the left, and scores for the four-year olds on the right.

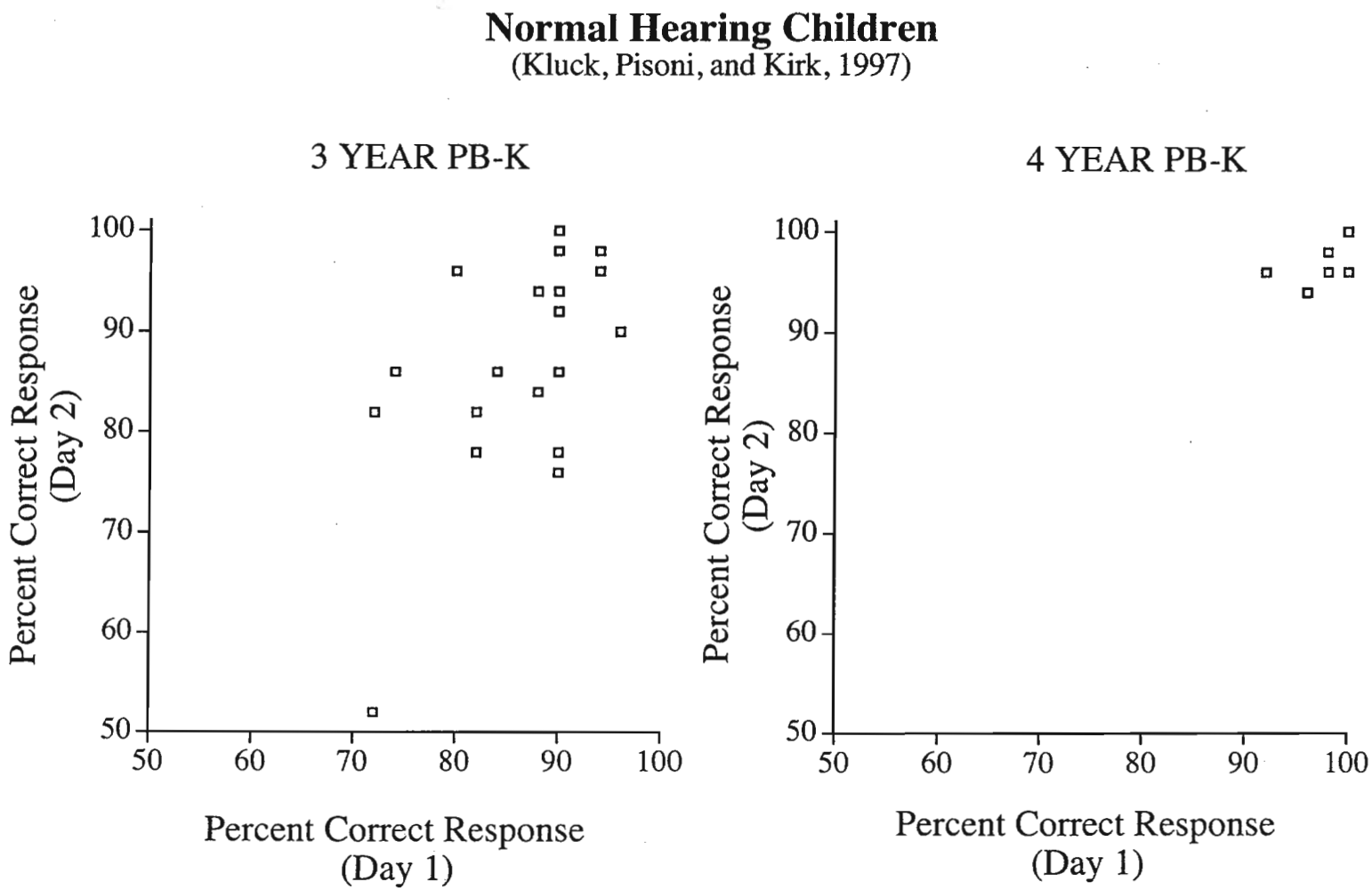

Figure 5. Scatterplots of percent correct scores for MLNT for Day 1 and Day 2, with scores for the three-year olds on the left, and scores for the four-year olds on the right.

Table 4

Parents familiarity ratings of their child's knowledge of words used in perception tests

	3 Year Olds [a]		4 Year Olds [b]	
	Mean (max. = 7)	(SD)	Mean (max. = 7)	(SD)
LNT FAM	6.74	.044	6.87	0.15
MLNT FAM	6.41	0.61	6.79	0.16
PBK FAM	5.95	1.0	6.58	0.37

Note. [a] N= 20 for these ratings. [b] N= 7 for these ratings.

The parents also completed a rating scale to assess their own vocabulary using the PFAM test. Four scores were obtained from the PFAM test: a low, medium, and high familiarity words, and an average of those three scores. Scores for both the three- and four- year olds are reported in Table 5.

Table 5

Means for PFAM test from both age groups

	PFAM HI	PFAM MED	PFAM LO	PFAM AVE
3 Year Olds	6.6	4.6	2.9	4.6
(SD) N=14	0.5	1.6	1.6	0.9
4 Year Olds	6.8	5.2	2.8	4.9
(SD) N=7	0.2	0.7	0.8	0.6

Discussion

Findings and Conclusions

The present investigation obtained several novel findings concerning the abilities of young children to recognize spoken words in isolation. The study was originally designed with two primary goals. First, we wanted to find the age at which a normal hearing child can successfully perform the LNT and MLNT tests. The results showed that normal hearing young children do very well on these three tasks, indicating that there may be another underlying reason, besides the difficulty of the vocabulary, that causes hearing

impaired children to perform poorly on these three open-set tests. We found that three- and four- year old normal hearing children can successfully perform the PBK, LNT and MLNT open-set word recognition tests at near ceiling levels of performance. These findings can be used as a benchmark to compare performance abilities of young hearing-impaired children to normal hearing children.

In general, both groups performed at ceiling levels on the word recognition tests. Performance by both age groups was very high, and there was very little variance in the results. This is the main reason why we were unable to establish test-retest reliability for the PBK, LNT, and MLNT.

The construction and original design of the LNT and MLNT were motivated by several factors. These tests were intended to be easier open-set tests than the older PBK, in order to raise performance levels by cochlear implant users. These new tests were also theoretically based and were designed to study the effects of lexical neighborhood density and word frequency on word recognition. Results from this study did not provide any information about lexical organization in normal hearing 3 and 4 year old children, because subjects performed extremely well, at close to ceiling levels, on both "easy" and "hard" lists.

Familiarity ratings completed by parents provided additional information, in terms of numerical ratings, about their child's vocabulary knowledge. These measures are easily obtained, because parents fill out the rating while the child participates in the experiment.

Scores on the LNT FAM and MLNT FAM for the three- year olds were higher than scores on the PBK FAM. This finding supports the hypothesis from Kirk et al. (1995) that the vocabulary on the LNT and MLNT tests is simpler that the vocabulary on the PBK test.

In summary, the findings from this study can be used as a benchmark for the LNT and MLNT, where performance of hearing-impaired children on these tests can be compared to the results from the present study. We now know that the LNT and MLNT are appropriate for three- and four- year old children, and these findings can be used to compare the performance of a hearing-impaired child to the high performance of a normal hearing child.

Directions for Future Research

There are several important directions for future research using these tests. Further studies with normal hearing children may be able to provide more information about lexical organization in this population.

Several explanations are possible for the high performance of normal hearing children on open-set speech perception tests. One explanation may be that they know and understand the meaning of the target word. Another is that they are able imitate the sound sequence of the target word. To further investigate open-set test performance in normal hearing children, a future study could be conducted using words and pseudo-words which are unfamiliar to young children. Such a study would be informative because we would learn the reasons why young children are able to repeat the words they hear. For example, are they successful because they are familiar with the syllable structure of English, or do they actually need to know the meaning of the target word in order reproduce it correctly?

Future studies may also provide further evidence concerning the cause of poor performance of pediatric cochlear implant users on open-set tests. Does poor performance result because children with

cochlear implants are unfamiliar with the vocabulary on the tests, or is there another more fundamental reason underlying their perceptual processes which are causing poor performance? Is it possible that the task of perceiving words in isolation, with no context, and then having to repeat isolated words from short term memory, is the cause of their low scores?

Finally, because the LNT and MLNT were designed to be used in testing pediatric cochlear implant users, perhaps it would be best to use this same population of subjects in establishing test-retest reliability for these new tests. It is most important that these tests are reliable when used for testing cochlear implant users, instead of with normal hearing children.

In summary, the present study of normal-hearing children has shown that this population can achieve high levels of performance on open-set word recognition tests. This finding, which may be used as an important bench mark to compare performance levels of various subject populations, encourages further investigation into the reasons behind the low levels of performance displayed by pediatric implant users. Study of normal hearing children may provide some insight into this problem. The results of the present study show that normal hearing three- year old children have little difficulty recognizing words from the PBK, LNT and MLNT tests.

The high scores obtained by normal hearing young children from the current study suggest that the poor performance of pediatric cochlear implant users may be due to other factors unrelated to the specific vocabulary items used on the open-set tests of word recognition. We believe it is important to try to identify some of these factors in future studies of children with cochlear implants.

References

Anastasi, A. *Psychological Testing 2nd Ed.* (1961). New York, NY: The Macmillan Company (p. 118).

Dunn, L. & Dunn, L. (1981). *Peabody Picture Vocabulary Test-Revised.* Circle Pines, MN: American Guidance Service.

Goldinger, S.D., Luce, P.A., & Pisoni, D.B., (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28,* 501-518.

Haskins, H. (1949). *A phonetically balanced test of speech discrimination for children.* Unpublished master's thesis, Northwestern University, Evanston, IL.

Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects of spoken word recognition by pediatric cochlear implant users. *Ear & Hearing, 16,* 470-481.

Kirk, K.I., Diefendorf, A.O., Pisoni, D.B., & Robbins, A.M. (1997). Assessing speech perception in children. In Mendel, L.L., & Danhauer, J.L. (Ed.), *Audiologic Evaluation and Management and Speech Perception Assessment.* San Diego, CA: Singular Publishing Group, Inc.

Logan, J.S. (1992). A computational analysis of young children's lexicons. *Research on Spoken Language Processing Technical Report No. 8.* Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P.A. (1986) Neighborhoods of words in the mental lexicon. *Research on Speech Perception Technical Report No. 6*. Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P.A., Pisoni, D.B., & Goldinger, S.D., (1990). Similarity neighborhoods of spoken words. In G.T.M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge, MA: MIT Press.

MacArthur Communicative Development Inventory: Words and Sentences. (1993). San Diego, CA: Singular Publishing Group, Inc.

MacWhinney, B., & Snow, C. (1985) The child language data exchange system. *Journal of Child Language, 12*, 271-296.

Nusbaum, H.C., Pisoni, D.B., Davis, C.K. (1984). Sizing Up the Hoosier Mental Lexicon: Measuring the Familiarity of 20,000 Words. *Research on Speech Perception Progress Report No. 10*. Bloomington, IN: Speech Research Laboratory, Indiana University.

Osberger, M.J., Miyamoto, R.T., Zimmerman-Phillips, S., Kemick, J.L., Stroer, B.S., Firszt, J.B., & Novak, M.A. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear and Hearing, 12*, 66S-80S.

Pisoni, D.B., Miyamoto, C., Kirk, K.I., Sommers, M.S., & Osberger, M.J. (1994). Sources of variability in speech perception by hearing-impaired listeners. Poster presented at the *17th Midwinter Research Meeting of the Association for Research in Otolaryngology*, St. Petersburg, FL., February 6-10, 1994.

Pisoni, D.B., Svirsky, M.A., Kirk, K.I., & Miyamoto, R.T. (1997). Looking at the "stars": A first report on the intercorrelations among measures of speech perception, intelligibility and language in pediatric cochlear implant users. Paper presented at the *5th International Cochlear Implant Conference,* New York, NY, May 1-3, 1997.

Rescorla, L. (1989). The language development survey: a screening tool for delayed language in toddlers. *American Speech-Language-Hearing Association, 54*, 587-599.