1998

Evaluating lexically controlled word lists for open set speech perception testing in profoundly hearing impaired children

Andrea Lynn Pfanstiel

Evaluating Lexically Controlled Word Lists for Open Set Speech Perception Testing in Profoundly Hearing Impaired Children

Independent Study
Andrea Lynn Pfanstiel

Supervised by Ann Geers, Ph.D. and Lisa S. Davidson, M.S. CCC-A
April 27, 1998

Evaluating Lexically Controlled Word Lists for Open Set Speech Recognition Testing in Profoundly Hearing Impaired Children

The Phonetically Balanced Kindergarten (PB-K) test is often used to assess the speech perception abilities of children who use multichannel cochlear implants. This test dates back to the 1940's when Haskins developed phonetically balanced (PB) lists composed of monosyllabic words selected from the vocabulary of kindergartners. The words were selected on the basis of the International Kindergarten Union vocabulary lists (Haskins, 1949). These lists were subsequently called the PB-K lists. Lists described as phonetically balanced are those in which the speech sounds in each reflect the relative occurrences of the phonemes in the American English-speaking population. The validity of using these phonetically balanced lists for assessing the speech perception abilities of hearing impaired children has been questioned for many reasons. First, results of PB-K testing may fail to show the spread of data expected from a group with diverse speech perception skills. Differences in their spoken word recognition abilities are due, in part, to age of onset of hearing loss, duration of hearing loss, degree of hearing loss, and length of sensory aid use. Some hearing impaired children, especially those with prelingual deafness, may achieve low word recognition scores because they are unfamiliar with the PB-K words. As a result, their scores may be lower than expected for their degree of hearing impairment.

Kirk, Pisoni, and Osberger (1995) reflected that teacher reports of children's speech perception abilities while using their multichannel cochlear implants often conflicted with the results obtained from speech perception testing using the PB-K list. That is, teachers reported better speech perception abilities in daily activities for certain children than their PB-K scores indicated. As a result, the use of the PB-K as a valid test of speech perception has been questioned.

The biggest criticism of the PB-K list is its vocabulary level. Word lists devised with the phonetically balanced criteria often contain words which are unfamiliar to profoundly deaf children. Sanderson-Leepa and Rintelman (1976) showed that preschoolers with normal hearing at age 3.5 obtained significantly lower PB-K scores than older children. As a result, clinicians should be careful in administering the PB-K test unless they are sure that the receptive vocabulary of the child approaches at least that of a kindergartner with normal hearing. However, deaf children typically demonstrate vocabulary levels that are about half their chronological age (Boothroyd, Geers, and Moog, 1991). If the word presented to the child is not within his receptive vocabulary, he may select a phonemically similar word which is contained in his vocabulary, or he may give no response to the PB-K word stimulus. As a result, the PB-K is not purely testing speech perception abilities, for the word lists were not designed to control for these lexical factors.

Given the PB-K's inherent problem of vocabulary, the use of new word lists specifically designed to control lexical factors has been examined. This is important for many reasons. In order to assess improvements in open set speech perception abilities of pediatric sensory aid users, a word recognition test sensitive to such changes is vital. However, to the extent that low PB-K scores are affected by vocabulary knowledge, results can provide little data regarding the efficacy of rehabilitation and auditory training. As a result, young deaf sensory aid users may appear incapable of open set speech perception when they really are. A new test sensitive to the differences between sensory aid users is quite important to better differentiate speech perception skills in this heterogeneous population.

To help control these lexical factors in speech perception testing, Kirk, Pisoni, and Osberger (1995) developed two new word lists to be used for open set speech perception testing. In an attempt to control the problems of vocabulary, they chose words based on frequency of occurrence from the Child Language Data Exchange System (CHILDES) database to be included in new word lists. This database contains transcripts of verbal exchanges between a child and caregiver or another child. Words from the utterances of children ages 3-5 years of age were analyzed for frequency of occurrence and lexical neighborhood. The median value of all monosyllable words (except proper nouns, possessives, contractions, plurals, and inflected forms of words) was determined for these two factors. Frequency refers to the number of occurrences of a given word within the transcripts analyzed. The median word frequency was four occurrences. Lexical neighbors are those words that differ by only one phoneme from the target word. An example is *juice* which has only *moose, loose* and *June* as lexical neighbors. C*ut* has some of the following lexical neighbors: *cat, come*, *but*, and *nut*. The median was four neighbors per target word. Two new 50-word lists were developed based on the factors of word frequency and lexical neighbors. One list consists of "easy" words, defined as those words above the median value for word frequency and below the median value for lexical neighbors. Words included on "hard" list were below the median value for word frequency and above the median value for lexical neighbors. These two new word lists comprise the Lexical Neighborhood Test (LNT). Kirk et al. administered the PB-K, the "easy" LNT test, and the "hard" LNT test using live voice presentation to a group of pediatric multichannel cochlear implant users. The percent words correct by the 19 prelingually deaf children ranged from 20%-72% correct ($\overline{X} = 29.6\%$) for the "easy" LNT words, 12%-72% correct ($\overline{X} = 23.4\%$) for the "hard" LNT list, and 4% to 54%

correct ($\overline{X} = 13.9\%$) on the PB-K test. Thus, they showed that the LNT tests produced overall higher scores and greater variability than was observed for the PB-K.

A second set of word lists, the Multisyllabic Lexical Neighborhood Test (MLNT) was also developed to assess how listeners use word length and syllable structure in word recognition tasks. The MLNT consists of an "easy" and a "hard" list each containing 24 two or three-syllable words. Words were chosen using the same procedures as in choosing words for the LNT lists except that monosyllabic words were excluded from consideration.

The purpose of this study was to replicate the experiment performed by Kirk, Pisoni, and Osberger but using pre-recorded stimulus presentation. If the results are similar to those found in the previous study, then this might indicate the LNT test is more clinically useful than the PB-K in estimating speech perception abilities in pediatric cochlear implant users.

Also, the researcher wished to examine the usefulness of the MLNT lists in assessing the speech perception abilities of pediatric sensory aid users. Perhaps, for children who perform poorly on the LNT lists, the MLNT is a better test for assessing their open set speech perception abilities. Because multisyllabic test items provide greater linguistic and contextual redundancy, scores would be expected to higher for multisyllabic than for monosyllabic stimuli.

Finally, a subset of 10 children were administered a 1/2 "easy" LNT list and a 1/2 "hard" LNT list using monitored live voice to examine the comparability of these scores and those obtained using pre-recorded stimuli.

EXPERIMENT ONE

The purpose of Experiment One was to determine whether or not differences in the lexical characteristics of word lists influence the word recognition scores of severe-to-profoundly deaf children who user hearing aids or cochlear implants, and to determine if lexically controlled word lists give a broader range of scores than phonetically balanced word lists.

METHOD

Subjects:

Subjects for Experiment One included 8 elementary school children from CID who used binaural hearing aids and 12 elementary school children who used the Nucleus 22 Channel Cochlear Implant utilizing the SPEAK speech coding strategy. The ages of the children ranged from 7 years, 0 months to 13 years, 11 months. All subjects demonstrated a pre-lingual severe to profound hearing loss. Specific requirements for all subjects used in the study included:

1.) All subjects scored above 75% correct on the monosyllable identification subtest of the Early Speech Perception Test (ESP),

2.) All subjects scored above 30% correct on the Word Identification by Picture Inventory (WIPI) test,

3.) All cochlear implant subjects were be implanted for a minimum of 1 year before the test date,

4.) For cochlear implant subjects, no additional hearing devices were used in the non-implanted ear.

Subject information is presented in Table 1.

Device Characteristics:

The 12 cochlear implant users wore the Spectra 22 speech processor programmed in the Spectral Peak (SPEAK) speech coding strategy. With this coding strategy, the processor scans the output of a bank of up to 20 filters that are used to analyze the incoming speech signal. The processor then selects, on average, the six filters with the strongest amplitude and activates the electrodes assigned to those filters. The eight hearing aid users wore conventional linear hearing aids on both ears.

Materials:

All testing was performed in a sound-treated chamber. Prior to testing, all words were recorded, digitized, and stored in a Macintosh computer using the Soundedit 16 program. A GSI 16 audiometer was used to deliver the digitized speech through a loudspeaker into the test room at 55 dBHL (normal conversational level). School files were used to access background information and past test scores for all subjects.

Procedure:

All 20 subjects (8 hearing aid users, 12 cochlear implant users) were administered a PB-K list, the "easy" LNT test, and the "hard" LNT list. Order of test presentation was counterbalanced across subjects. The words were presented at a conversational level (55 dBHL) using a pre-recorded voice delivered through a Macintosh computer. Subjects were required to repeat each word after it was presented. Each word was presented only once. Subjects' responses were recorded on paper and judged as either correct or incorrect by the examiner. Word correct scores were then computed for each word list. Table 2 lists the PB-K test, "easy" LNT list, and the "hard" LNT list used in testing.

Results:

Figure 1 presents the mean percent of words correctly identified on each monosyllabic word list. The percent of words correctly identified by the 20 subjects ranged from 0% to 34% correct on the LNT "easy" word list, from 0% to 36% correct on the LNT "hard" list, and from 0% to 16% correct on the PB-K test. Only six of the twenty subjects showed a decrease in the word recognition score on the LNT "hard" list as compared to the LNT "easy" list. Three subjects had identical scores, and 11 subjects showed an improvement in word recognition scores on the LNT "hard" list as compared to the LNT "easy" list. All subjects showed an improvement in word recognition scores using either LNT list as compared to using the PB-K test.

Discussion:

The results of experiment one suggest that word lists specifically designed to contain stimulus items expected to be in a deaf child's vocabulary do yield higher word recognition scores than do word lists designed to contain phonetically balanced stimulus items. That is, the mean scores for both the "easy" and "hard" LNT lists (15.4% and 15.7% respectively) were better than the mean score for the PB-K test (4.2%). Also, the standard deviations for the "easy" and "hard" LNT lists (11.51% and 12.5% respectively) were higher than the standard deviation for the PB-K test (5.14%). This suggests that using the LNT lists in testing would give a broader range of scores than would the PB-K test. Presumably, this would lead to a better correlation between a particular child's word recognition score and parental and/or teacher reports of speech perception abilities at home or in the classroom. That is, a child with reportedly excellent speech perception abilities would be expected to score better when tested using the LNT list than a child with reportedly poor speech perception abilities. A moderately high correlation (.53) was obtained for a sample of 46 implanted children tested in the CID "Education and Implants" study (Geers, 1998) between LNT scores and parental ratings on the "Auditory Responsiveness" questionnaire. The LNT lists would then be expected to be sensitive to changes in speech perception abilities over time.

The lack of a significant difference found in word recognition scores between the "easy" and "hard" LNT lists was unexpected. Presumably, children would be expected to score better on the "easy" LNT list as this list was designed to contain more familiar words and words with fewer lexical neighbors than the "hard" LNT list. This suggests that these children do not organize words into similarity neighborhoods in their long-term memories as described by Kirk et al. According to the results of this study, however, the number of lexical neighbors does not appear to have an effect on word recognition scores. The children tested did not seem to recognize words in the context of other words as was expected. This observation, however is different from that of Kirk et. al. (1995) in which their pediatric subjects performed better on the LNT "easy" list than on the LNT "hard" list, suggesting that pediatric sensory device users do recognize words in the context of other words and that words are organized into similarity neighborhoods in their long-term memories. The difference may be due, in part, to the lower scores obtained by the experimental group of the study under question (15%) compared with Kirk et al.'s group who averaged better than 30% words correct. At an average of only 15% words correct, the CID group does not appear to have strong enough skills to "know" many lexical neighbors. As a result, the "easy" and "hard" lists become equivalent. In the CID "Education and Implants" study (Geers, 1998), in which eight and nine-year old children with more than three years of implant use were tested on both the "easy" and "hard" LNT lists, a significant difference was observed. However, scores were generally higher than those observed in this study (43% correct for "easy" LNT and 38% correct for "hard" LNT).

EXPERIMENT TWO

The purpose of Experiment Two was to determine if the use of multisyllabic test items would yield higher word recognition scores than would monosyllabic test items. Presumably, scores using monosyllabic stimuli would yield higher scores because these items are less easily confused with other words than are monosyllabic stimuli. Also, monosyllabic words are particularly difficult because the redundant linguistic and contextual cues available in multisyllabic words and sentences are absent in monosyllabic words. Another purpose was to compare the scores elicited using the pre-recorded stimuli in experiment one to those elicited using monitored live voice testing. Kirk et al.'s data were collected live voice, which may have accounted for the substantially higher scores than those found by this researcher in experiment one.

METHOD

Subjects:

Subjects for Experiment Two included 3 elementary school children from CID who used binaural hearing aids and 7 elementary school children who used the Nucleus 22 Channel Cochlear Implant utilizing the SPEAK speech coding strategy. The ages of the children ranged from 7 years, 0 months to 12 years, 10 months. All subjects demonstrated a pre-lingual severe to profound hearing loss. Specific requirements for all subjects used in the study were the same as used in Experiment One. Subject information is presented in Table 3.

Device Characteristics:

The 7 cochlear implant users used the Spectra 22 speech processor programmed in the Spectral Peak (SPEAK) speech coding strategy. The three hearing aid users wore conventional linear hearing aids on both ears.

Materials:

All testing was performed in a sound-treated chamber. A GSI 16 audiometer was used to deliver the monitored live voice words through a loudspeaker into the test room at 55 dBHL (normal conversational level). Each word was spoken so that the VU meter peaked at 0 dB. School files were used to access background information and past test scores for all subjects.

Procedure:

All 10 subjects (3 hearing aid users, 7 cochlear implant users) were administered a 25-item "easy" LNT list, a 25-item "hard" LNT list, a 24-item "easy" MLNT list, and a 24-item "hard" MLNT list. Order of test presentation was counterbalanced across subjects. Subjects were required to repeat each word after it was presented. Each word was presented only once. Subjects' responses were recorded on paper and judged as either correct or incorrect by the examiner. Word correct scores were then computed for each word list. Table 4 lists the stimuli used in testing.

Results:

Figure 2 presents the mean percent of words correctly identified on each word list. Percent word correct scores ranged from 20.84% to 79.16% for the "easy" MLNT list, from 20.84% to 75% for the "hard" MLNT list, from 20% to 68% for the LNT "easy" list, and from 8% to 52% for the LNT "hard" list. Word correct scores were highest when using the "easy" MLNT list ($\overline{X}$ = 60.0%), were essentially the same when using the "hard" MLNT list ($\overline{X}$ = 46.54%) or the "easy" LNT list ($\overline{X}$ = 46.8%), and were the poorest when using the "hard" LNT list ($\overline{X}$ = 35.6%). On average, multisyllabic word recognition performance was better than monosyllabic word recognition for both "easy" and "hard" word lists. The difference between "easy" and "hard" word lists, absent in experiment one using recorded word lists were apparent for both the LNT and the MLNT word lists. Results are similar to those reported by Kirk et al.

The use of monitored live voice testing did yield better word recognition scores than did recorded testing. Table 5 lists the average scores for "easy" and "hard" lists in both live voice and recorded conditions along with the average improvement observed when using monitored live voice testing. The range of improvement for the "easy" LNT list was 18% to 44% and for the "hard" list, the range of improvement was 8% to 26%. Thus, the effect of using recorded stimulus presentation was a dramatic reduction in open set word recognition performance.

Discussion:

The results of experiment two suggest that pediatric sensory aid users use word length to assist them in word recognition. Presumably, this is because multisyllabic words have fewer lexical neighbors than monosyllabic words. Also, the multisyllabic words have greater redundancy of linguistic and contextual cues which the pediatric sensory aid users appear to utilize in word recognition. This finding is in agreement with other research with listeners with normal hearing (Cluff and Luce, 1990). These results suggest that the MLNT lists may be useful in testing the open set speech perception abilities of pediatric sensory aid users who perform poorly on even the LNT word lists. The MLNT lists may be particularly useful in assessing the speech perception abilities of children with limited auditory perception skills.

In contrast to Experiment One, scores were higher in the live voice condition for the word lists labeled "easy" compared to those labeled "hard." That is, words that were high in frequency and low in neighborhood density were identified with greater accuracy than were words that were low in frequency and high in neighborhood density. This observation is likely related to the overall higher scores evident in this live voice condition. More phonemic information was being transmitted, and thus potentially "confused" than in the recorded condition.

The average improvement was 28% for the "easy" LNT list and 18.4% for the "hard" LNT list when using monitored live voice testing over recorded stimuli. Pre-recorded testing is extremely important for preserving consistency in presentation from one patient to another or for repeated assessment of the same individual. The often cited advantage of using monitored live voice testing is the flexibility this method affords to the examiner. However, with the advent of the use of computers in speech testing, the same amount of flexibility can be maintained as words are presented only when the examiner instructs the computer to do so. Because it is extremely difficult for any examiner to present a consistent signal from one listener to the next or from one test session of the same individual to the next, it is advised that only pre-recorded materials be used in research. However, the effects of recorded stimulus presentation on open set speech recognition scores requires further documentation.

Table 1

Ages and Length of CI Use

Subject	C.A. @ Test	Length of CI Use	PPVT Age Score	Better Ear SFA
1)	12 yrs, 10 mos	9 yrs, 10 mos	4 yrs, 10 mos	120 dB
2)	11 yrs, 0 mos	4 yrs, 3 mos	5 yrs, 2 mos	103 dB
3)	12 yrs, 8 mos	5 yrs, 6 mos	4 yrs, 1 mos	115 dB
4)	9 yrs, 2 mos	6 yrs, 0 mos	4 yrs, 7 mos	113 dB
5)	10 yrs, 4 mos	5 yrs, 4 mos	4 yrs, 10 mos	118 dB
6)	10 yrs, 4 mos	6 yrs, 4 mos	4 yrs, 4 mos	116 dB
7)	10 yrs, 7 mos	3 yrs, 11mos	5 yrs, 9 mos	96 dB
8)	10 yrs, 4 mos	4 yrs, 4 mos	4 yrs, 9 mos	120 dB
9)	11 yrs, 2 mos	7 yrs, 4 mos	5 yrs, 1 mos	113 dB
10)	7 yrs, 9 mos	4 yrs, 8 mos	3 yrs, 1 mos	115 dB
11)	7 yrs, 10 mos	4 yrs, 9 mos	5 yrs, 2 mos	120 dB
12)	7 yrs, 0 mos	3 yrs, 0 mos	4 yrs, 1 mos	120 dB
$\overline{X}$ =	10 yrs, 1 mos	5 yrs, 5 mos	4 yrs, 8 mos	114 dB
standard deviation=	1 yr, 9 mos	1yr, 9 mos	0 yrs, 8 mos	7 dB
13)	13 yrs, 11 mos	N/A	4 yrs, 1 mos	98 dB
14)	10 yrs, 4 mos	N/A	4 yrs, 10 mos	98 dB
15)	8 yrs, 4 mos	N/A	5 yrs, 9 mos	98 dB
16)	9 yrs, 8 mos	N/A	5 yrs, 1 mos	101 dB
17)	10 yrs, 5 mos	N/A	4 yrs, 8 mos	96 dB
18)	8 yrs, 3 mos	N/A	4 yrs, 6 mos	100 dB
19)	11 yrs, 6 mos	N/A	5 yrs, 2 mos	98 dB
20)	8 yrs, 2 mos	N/A	4 yrs, 3 mos	95 dB
$\overline{X}$ =	10 yrs, 9 mos		4 yrs, 9 mos	98 dB
standard deviation=	1 yr, 10 mos		0 yrs, 6 mos	1.8 dB

Table 2

Stimuli Used-Experiment One

"Easy" LNT		"Hard" LNT		PB-K	
down	juice	ear	thumb	please	smile
truck	good	hand	pie	great	bath
mouth	drive	dry	wet	sled	slip
pig	time	zoo	fight	pants	ride
give	hard	goat	toe	rat	end
school	gray	toy	cut	bad	pink
boy	foot	call	pink	pinch	thank
put	orange	sing	hi	such	take
three	count	cut	song	bus	cart
farm	brown	wrong	fun	need	scab
fish	home	bed	use	ways	lay
green	old	fat	mine	five	class
catch	watch	man	ball	mouth	me
break	need	run	kick	rag	dish
house	food	hot	tea	put	neck
sit	dance	read	book	fed	beef
friend	live	grow	bone	fold	few
jump	stand	bag	work	hunt	use
bird	six	cake	dad	no	did
swim	cold	seat	game	box	hit
hold	push	nine	lost	are	pond
want	stop	sun	cook	teach	hot
snake	girl	bath	gum	slice	own
more	hurt	ten	cap	is	bead
white	cow	ride	meat	tree	shop

Table 3

Ages and Length of CI Use

Subject	C.A. @ Test	Length of CI Use	Better SFA	PPVT
1)	10 yrs, 4 mos	5 yrs, 4 mos	118 dB	4 yrs, 10 mos
2)	12 yrs, 8 mos	5 yrs, 6 mos	115 dB	4 yrs, 1 mos
3)	9 yrs, 2 mos	6 yrs, 0 mos	113 dB	4 yrs, 7 mos
4)	10 yrs, 4 mos	6 yrs, 4 mos	116 dB	4 yrs, 4 mos
5)	12 yrs, 10 mos	9 yrs, 10 mos	120 dB	4 yrs, 10 mos
6)	11 yrs, 0 mos	4 yrs, 3 mos	103 dB	5 yrs, 2 mos
7)	7 yrs, 0 mos	3 yrs, 0 mos	120 dB	4 yrs, 1 mos
8)	8 yrs, 3 mos	N/A	100 dB	4 yrs, 6 mos
9)	11 yrs, 6 mos	N/A	98 dB	5 yrs, 2 mos
10)	8 yrs, 2 mos	N/A	95 dB	4 yrs, 3 mos
$\overline{X}$ =	10 yrs, 1mos	5 yrs, 9 mos	109.8 dB	4 yrs, 7 mos
Standard deviation=	1 yr, 10 mos	1 yr, 1mos	9.2 dB	0 yrs, 5 mos

Table 4

Stimuli Used-Experiment Two

"Easy" LNT	"Hard" LNT	"Easy" MLNT	"Hard" MLNT
juice	thumb	children	butter
good	pie	animal	lion
drive	wet	monkey	money
time	fight	finger	jelly
hard	toe	pocket	yellow
gray	cut	apple	purple
foot	pink	morning	hello
orange	hi	sugar	carry
count	song	alright	corner
brown	fun	about	heaven
home	use	because	measles
old	mine	crazy	ocean
watch	ball	water	puppy
need	kick	banana	pickle
food	tea	glasses	button
dance	book	airplane	summer
live	bone	window	bottom
stand	work	tiger	finish
six	dad	cookie	bunny
cold	game	again	belly
push	lost	another	couple
stop	cook	almost	under
girl	gum	broken	naughty
hurt	cap	china	really
cow	meat		

Table 5

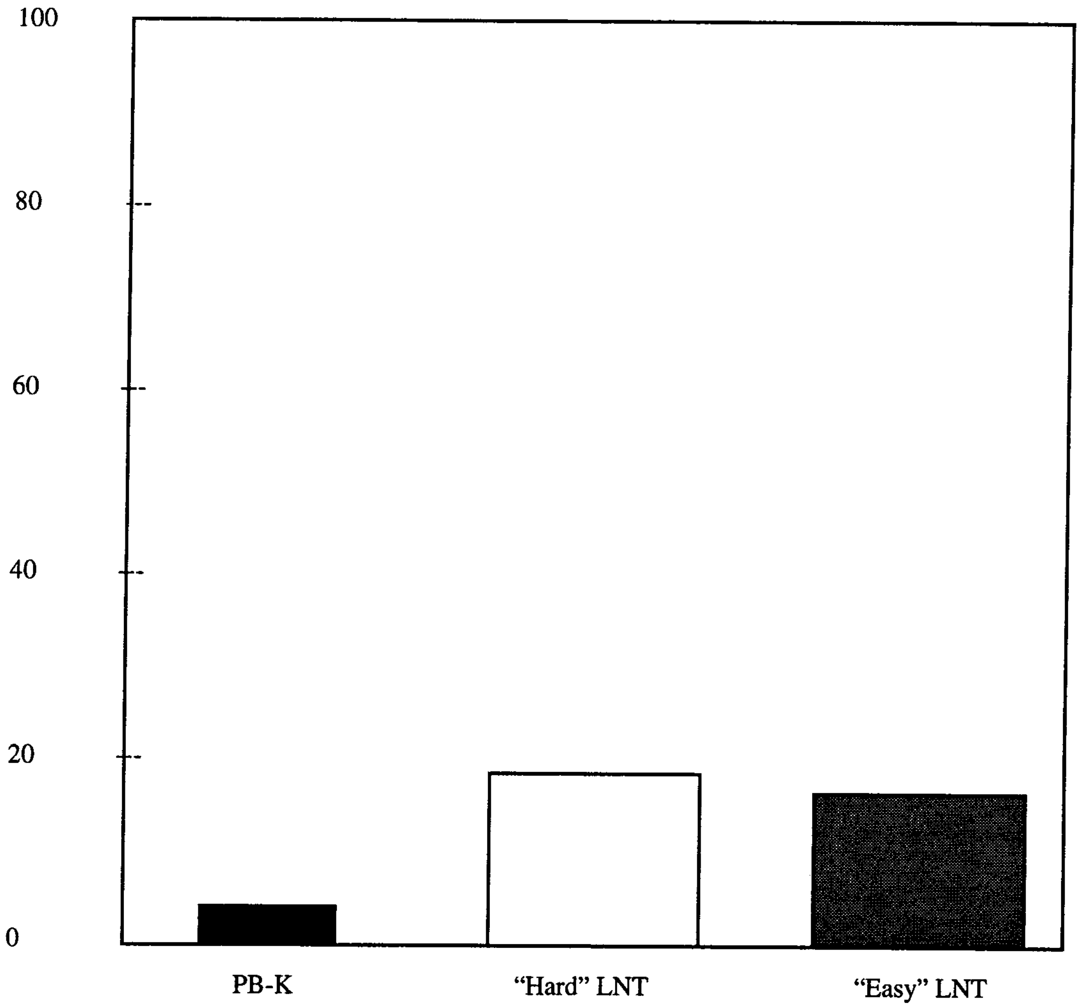
Average Improvements with MLV Testing

AVERAGE

Test	Live Voice	Recorded	Average Improvement	Range	standard deviation
"Easy" LNT	46.8%	15.4%	28.0%	18-44%	9.8%
"Hard" LNT	35.6%	15.7%	18.4%	8-26%	5.8%

Figure 1

Word Correct Scores-Experiment One



PB-K	"Hard" LNT	"Easy" LNT
$\overline{X} = 4.2\%$	$\overline{X} = 15.7\%$	$\overline{X} = 15.4\%$
standard deviation= 5.14%	standard deviation= 12.5%	standard deviation= 11.51%
Range= 0-16%	Range= 0-30%	Range= 0-34%

Figure 2

Word Correct Scores-Experiment Two