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

Some Computational Analyses of the PBK Test: Effects of Frequency and Lexical Density on Spoken Word Recognition[1]

Ted A. Meyer[2] and David B. Pisoni

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

[1] This study was supported by NIH/NIDCD Grants DC00064, DC00111, and NIH Training Grant DC00012. Portions of the paper were presented at the 21st Annual Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL, February, 1998. We would like to thank Darla J. Sallee and Linette A. Caldwell for clerical assistance and Dr. Steven B. Chin for comments and suggestions on an earlier version of the manuscript.
[2] Department of Otolaryngology–Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Some Computational Analyses of the PBK Test: Effects of Frequency and Lexical Density on Spoken Word Recognition

Abstract. The Phonetically Balanced Kindergarten (PBK) Test (Haskins, 1949) has been used for almost 50 years to assess spoken word recognition performance in children with hearing impairments. The test originally consisted of four lists of 50 words, but only three of the lists (Lists 1, 3 & 4) were considered "equivalent" enough to be used clinically with children. Our goal was to determine if the lexical properties of the different PBK lists could explain any differences between the three "equivalent" lists and the fourth PBK list (List 2) that has not been used in clinical testing. Word frequency and lexical neighborhood frequency and density measures were obtained from a computerized database for all of the words on the four lists from the PBK Test as well as the words from a single PB-50 (Egan, 1948) word list. The words in the "easy" PBK list (List 2) were of higher frequency than the words in the three "equivalent" lists. Moreover, the lexical neighborhoods of the words on the "easy" list contained fewer phonetically similar words than the neighborhoods of the words on the other three "equivalent" lists. The present computational analyses show that both word frequency and lexical neighborhood density influence the probability of correct word recognition in open-set speech intelligibility tests. The results of this computational analysis of the PBK Test provide additional support for the proposal that spoken words are recognized "relationally" in the context of other phonetically similar words in the lexicon. Implications of using open-set word recognition tests with children with hearing impairments are discussed with regard to the specific vocabulary and information processing demands of the PBK Test.

Introduction

The PBK Test is an open-set test of spoken word recognition that has been widely used over the years in clinical audiology to measure speech perception skills in young children, especially deaf children with cochlear implants (CIs) and hearing aids (Carney et al., 1991; Fryauf-Bertschy, Tyler, Kelsay, & Gantz, 1992; Fryauf-Bertschy, Tyler, Kelsay, Gantz, & Woodworth, 1997; Gantz, Tyler, Tye-Murray, & Fryauf-Bertschy, 1994; Gantz, Tyler, Woodworth, Tye-Murray, & Fryauf-Bertschy 1994; Kirk, Osberger, & Pisoni, 1995; Miyamoto, Osberger, Robbins, Myres, & Kessler 1993; Miyamoto et al., 1994; Osberger et al., 1991; Pisoni, Svirsky, Kirk, & Miyamoto, submitted; Staller, Beiter, Brimacombe, Mecklenburg, & Arndt, 1991; Staller, Dowell, Beiter, & Brimacombe, 1991; Waltzman et al., 1994; Waltzman et al., 1995). The test was originally developed by Harriet Haskins in 1949 as part of her Master's thesis at Northwestern University to fill a need for an open set test of speech perception that could be used for children of kindergarten age. Although Haskins' thesis was never published, three of her original four word lists are routinely used clinically and have been employed extensively over the years to measure speech perception performance in young children with hearing losses.

As stated above, results from the PBK Test have come to play an extremely important role in studies of speech perception of children with CIs. Performance on the PBK Test is often used as the primary and "defining" criterion to identify exceptionally good users of CIs, the so-called "Star" performers (Pisoni et al., submitted). Children who display high levels of performance on the PBK Test also tend to perform well on other perceptual tests that are routinely administered as part of a standard assessment battery such as the Minimal Pairs Test (Robbins, Renshaw, Miyamoto, Osberger & Pope, 1988), Common Phrases Test (Osberger et al., 1991), Lexical Neighborhood Test (LNT) (Kirk et al., 1995), Peabody Picture Vocabulary Test (Dunn & Dunn, 1981), and the Reynell Developmental Language Scales (Reynell & Gruber, 1990). In addition, and perhaps even more interesting, recent analyses have shown that children who score exceptionally well on the PBK Test also display very good speech intelligibility as measured by transcription scores of naive listeners (Pisoni et al., submitted; Svirsky, 1996). This finding suggests that these particular deaf children have not only acquired the perceptual skills needed to recognize isolated words but have also developed the means to encode, represent and retrieve the sound patterns of the spoken words in memory. These children have acquired control over those aspects of expressive language that are needed to access motor patterns in speech production to produce intelligible speech. We believe these are important milestones in speech perception and language development, and they deserve more detailed examination in order to understand the basis for these achievements in deaf children with CIs.

Researchers and clinicians alike have routinely observed enormous amounts of variability among users of CIs on all of the standard assessment instruments (Carney et al., 1991; Fryauf-Bertschy et al., 1992; Fryauf-Bertschy et al., 1997; Gantz, Tyler, Woodworth et al., 1994; Staller, Dowell et al., 1991; Staller, Beiter et al., 1991). In the case of the PBK Test, the range of performance for children covers almost the full scale from zero to approximately 90 percent correct word recognition in some cases (Carney et al., 1991; Fryauf-Bertschy et al., 1992; Fryauf-Bertschy et al., 1997; Gantz, Tyler, Woodworth et al., 1994; Osberger et al., 1991; Waltzman et al., 1994). Upon closer inspection, the individual scores on the PBK Test tend to follow a bimodal distribution of performance: some of the children with CIs are able to perform well on the PBK open-set test whereas other children routinely score zero on the test (Lane, 1995). Why is there so much variability among CI users on the PBK Test? Why is the PBK Test apparently so difficult for some deaf children with CIs but much more manageable for other children? These are several of the questions we hope to answer in this paper.

In examining the task demands of the PBK Test, two major factors come to mind that are worth considering in some greater detail. First, we consider the specific vocabulary items used on this test. Another speech perception test for young deaf children was developed by Quick (1949) at the approximate time Haskins developed the PBK Test. Quick's closed-set test consisted of two lists of 25 monosyllabic words, with each word being accompanied by two very similar-sounding words. All of the English speech sounds were represented in the two lists, but the two lists did not maintain the phonetic balance of the Phonetically Balanced (PB) lists that were developed earlier in the Psycho-Acoustic Laboratory at Harvard University (Egan, 1948). Haskins originally developed the PBK lists from words taken from the well-known PB-50 lists for the adult listeners, but also used words that were in the 2500 words of highest frequency of words spoken by preschool children (The International Kindergarten Union, 1928). It has been suggested recently by Kirk et al. (1995) that children with limited vocabulary skills, such as very young children or children with CIs, may score poorly on or be unable to do the PBK Test at all because the specific words used on the test are simply too difficult and are not within the vocabulary of these children. Is this explanation of the difficulty of the PBK Test correct? How can this account be assessed?

At the present time, there are no vocabulary norms or lexical databases for deaf children, so it is difficult to compare the words on the PBK Test to words from the vocabularies of deaf children. Kirk et al. (1995) attempted a computational analysis of the words on the PBK Test using a lexical database constructed by Logan (1992). Logan's database contained samples of the vocabulary from normal-hearing children that was available in the CHILDES (Child Language Data Exchange System) database (MacWhinney & Snow, 1985). Kirk et al., however, found that a large number (approximately 69%) of the words on the PBK Test were, in fact, not within Logan's database. Thus, it is possible that the children with CIs cannot recognize the words on the PBK Test simply because they are not familiar with the specific items used in this test?

In addition to the specific vocabulary items used on the PBK Test, the lexical properties of the words themselves may also play an important role in recognition. Some words may be hard to perceive because the children simply do not know them. Alternatively, some words may be hard to perceive because they are phonetically confusable with many other similar-sounding words (see Luce & Pisoni, in press). And, some words may be hard to perceive because deaf children have difficulty making fine phonetic discriminations that are needed for identification of these particular sound patterns. In their computational analyses of the vocabulary of the PBK Test, Kirk et al. (1995) found that a large number of words on the PBK Test were perceptually difficult; that is, the words came from regions of the lexicon where there were many other phonetically-similar words that were higher in frequency than the target word. These observations suggest that discriminability and competition among lexical candidates that are phonetically similar to each other may also affect performance on the PBK test and may be another important factor that makes the words on this test difficult for deaf children with CIs (see Luce & Pisoni, in press). Thus, the PBK Test may be a very difficult test for young deaf children not only because many of the words are unfamiliar and therefore not part of their lexicon but also because many of the words are extremely difficult to perceive in isolation where the only available context is the presence of other phonetically-similar words in the child's language.

In addition to the specific vocabulary items used on the PBK Test, it is also possible that the specific task demands of an open-set word recognition test also affect performance, particularly in young deaf children who may not be able to make fine phonetic discriminations among different sound patterns and encode these into memory. An open-set word recognition test like the PBK Test has no external context or response constraints that a listener can rely on other than the knowledge of sound patterns and regularity of words in their lexicon. Recognizing a word in isolation therefore requires that the listener encode the sound pattern into working memory, and then access a motor program for the word from information stored in the lexicon in order to produce an utterance on demand in a repetition task. Children may therefore have difficulty with an open-set test not only because the vocabulary is unfamiliar but because they are not able to encode, represent or access the sound pattern of these novel items from memory in what is essentially an imitation task. Either or both of these alternatives are possible accounts of why the PBK Test is difficult for young deaf children. Because of the importance placed on open-set word recognition tests as a criterion for evaluating performance with CIs, we believe that it is critical at this time to gain an understanding of the reasons for the generally poor performance observed on the PBK test with hearing-impaired children and to analyze the specific task demands of this test.

After reading the original version of Harriet Haskins' thesis, we discovered several interesting findings that we wish to report here both for historical reasons, because her thesis was never published in a scientific journal and therefore has had only a very limited distribution within the hearing and speech science community, and for scientific reasons, because her original results obtained almost 50 years ago are still informative and bear on several current theoretical issues concerning speech perception and spoken word recognition in young children. In particular, based on our analyses reported below, we believe it is possible to provide a principled explanation of why the PBK Test is a difficult test for some children with CIs. Several new computational analyses of the specific words used on the PBK Test were conducted, and these results provide some new insights into the primary factors that influence word recognition in open-set tests. In the sections below, we first describe how the PBK word lists were constructed and summarize Haskins's findings. Then we report the results of our computational analyses of the word lists. Finally, we propose an explanation of the pattern of her results and discuss the implications of these new findings for assessing speech perception and spoken word recognition in deaf children with CIs.

Although not widely known, Haskins actually constructed *four* separate lists of PBK words and then collected speech intelligibility data on these materials which were included in her thesis. However, only three of these original lists are currently being used by audiologists in the clinic and in research laboratories. Apparently, the fourth PBK list was never used because, according to Haskins, her speech intelligibility scores with adult listeners with normal hearing showed that the words on this particular list (PBK List 2) were "more audible" that the words used on the other three lists. In order to identify the basis for these differences, we have carried out a series of "computational" analyses of the words used in all four of Haskins' word lists and we report the results below. We believe these new findings on the words used in the PBK Test are important and should be of interest to researchers and clinicians who use speech intelligibility tests to measure and assess changes in speech perception and word recognition skills in young children, particularly young deaf children with hearing aids or CIs.

As mentioned previously, Haskins collected speech intelligibility data with these four word lists using normal-hearing adults as subjects. To the best of our knowledge, no speech intelligibility data were ever collected from young children with normal hearing, the target population that the materials were originally developed for at the time, nor have any speech intelligibility data been published from children with normal hearing using these lists. Because of the important role that the PBK Test has played in recent years in assessments of speech perception and word recognition skills in young children, we critically examine both issues below.

In the first section of this paper, we report new computational analyses on all four of Haskins' original word lists. The results of these analyses permit us to offer a theoretically-motivated explanation for the differences in speech intelligibility she observed between the lists of words with adult listeners. Not only can we now account for differences among the four original lists, but these new computational analyses of Haskins's data also provide additional support for a recent model of spoken word recognition -- the Neighborhood Activation Model (NAM) (Luce, 1986; Luce & Pisoni, in press; Luce, Pisoni & Goldinger, 1990). More importantly, the results of these analyses provide new evidence for the proposal that spoken words are recognized "relationally" in the context of other phonetically similar words in the listener's lexicon. Using the Neighborhood Activation Model of spoken word recognition as our theoretical framework, we are able to offer an explanation for why some words are easy to recognize and why other words are more difficult to recognize in an open-set test format using a few simple principles that characterize the recognition process for spoken words (see Luce & Pisoni, in press). Finally, we offer some comments about the implications of these findings for assessment of speech perception in deaf children.

Methods

Word Lists

Haskins recorded a single randomization of the four lists of the PBK Test. She then presented the four PBK lists and one of the PB-50 lists (List 13) at sequentially higher signal levels to adult listeners with normal hearing. The highest level was approximately 27 dB SL (0 dB SL was determined by having the subjects set an attenuator to a level at which they felt they were obtaining approximately 50% of a sample of continuous discourse). Listeners were asked to identify the words using an open-set format. Haskins found that List 2 was "easier" than the other lists at attenuation levels of 50 (12 dB SL) and 60 (2 dB SL) dB, but at the highest level (27 dB SL) the scores from all five lists were essentially equivalent, due to a ceiling effect. Haskins stated that she felt that List 2 was "easier" than the other lists at all levels of presentation, although a significant difference between the lists emerged only at 50 and 60 dB attenuation. She also stated that an "item analysis" of the individual words in PBK List 2 might yield additional information on ways to improve the list, or to make it more like the other lists, but she did not indicate what kind of item analysis to do or what perceptual dimensions of spoken words might be relevant to speech intelligibility performance or to the differences in performance she observed.

Lexical Neighborhood Database

The four PBK lists of Haskins (1949) and List 13 of the PB-50 words (Egan, 1948) were analyzed using several techniques to compute similarity spaces for spoken words (see Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). A computational analysis of the specific words used on all five lists was carried out using a computerized version of a 20,000-word Webster's Pocket Dictionary (Pisoni et al., 1985). The word lists are given in Table 1. From the pocket dictionary, we obtained word frequencies (Kucera & Francis, 1967), as well as lexical neighborhood frequencies and densities. Word frequency refers to the frequency counts from the Kucera and Francis (1967) norms. A "lexical neighborhood" of a word is defined as all words in the pocket dictionary that differ from the stimulus word by a single phoneme *substitution, addition,* or *deletion* (Greenberg & Jenkins, 1963). For example, if the stimulus is "pit", an example of a neighbor by *substitution* is "bit", by *deletion* "it", and by *addition* "spit". Neighborhood density refers to the number of words in the lexical neighborhood, and neighborhood frequency refers to the frequency counts (Kucera and Francis, 1967) of the words in the lexical neighborhood. The mean word frequency was calculated for all the words in each list. Because a few of the words had very high frequencies (e.g., "and" in List 2), the median word frequency was also calculated for each list. The median as a measure of central tendency tends to be less dependent upon outliers in asymmetrical or skewed distributions. Mean and median lexical neighborhood frequencies and densities were also calculated for the words in each individual list. Two additional measures were calculated for each word, the ratio of the word frequency to the neighborhood frequency and the ratio of the word frequency to the neighborhood density. These "second-order statistics" provide measures of the *relation* of a specific word to its neighbors and are used to quantify the amount of lexical competition among phonetically similar words in lexical memory.

Results

The results of the computational analyses for the four PBK lists (1-4) and the single PB-50 list (13) are shown in Tables 2 and 3. Haskins (1949) found that the words on List 2 were easier to recognize at several S/N than the words on the other lists, and she concluded that the items on List 2 were "more audible" than the words on the other three PBK lists at two intensity levels. Haskins obtained percent correct scores from 22 normal-hearing adults for the PBK words. The word lists are presented in Table 1, and the results from these intelligibility tests are shown in Table 2 (adapted from Haskins, 1949, Table VI). According to Haskins, at the middle intensity tested (2.2 dB SL), List 2 is significantly easier than Lists 1, 3, and 13. At the next higher intensity tested (12.2 dB SL), List 2 is significantly easier than all the other lists. At the highest intensity tested (27.2 dB SL), List 2 is significantly easier than only List 13, but at this level, performance is nearly perfect, and any differences between lists are confounded by ceiling effects.

Table 1.

Words on PBK Lists 1-4 and PB-50 List 13.

	PBK-1	PBK-2	PBK-3	PBK-4	PB-13
1	please	this	laugh	tire	bat
2	great	ma	falls	seed	beau
3	sled	pick	paste	purse	change
4	pants	glove	plow	quick	climb
5	rat	gun	page	room	corn
6	bad	forth	weed	bug	curb
7	pinch	trade	gray	that	deaf
8	such	each	park	sell	dog
9	bus	ask	wait	low	elk
10	need	wake	fat	rich	elm
11	ways	calf	ax	those	few
12	five	rope	cage	ache	fill
13	mouth	night	knife	black	fold
14	rag	chew	turn	else	for
15	put	guess	grab	nest	gem
16	fed	wave	rose	jay	grape
17	fold	cloud	lip	raw	grave
18	hunt	good	bee	true	hack
19	no	barn	bet	had	hate
20	box	left	his	cost	hook
21	are	shoe	sing	vase	jig
22	teach	flag	all	press	made
23	slice	rode	bless	fit	mood
24	is	hook	suit	bounce	mop
25	tree	front	splash	wide	moth
26	smile	toe	path	most	muff
27	bath	south	feed	thick	mush
28	slip	rest	next	if	my
29	ride	tongue	wreck	them	nag
30	end	best	waste	sheep	nice
31	pink	reach	crab	air	nip
32	thank	slide	peg	set	ought
33	take	food	freeze	dad	owe
34	cart	new	race	ship	patch
35	scab	ball	bud	case	pelt
36	lay	three	darn	you	plead
37	class	closed	fair	may	price
38	me	kept	sack	choose	pug
39	dish	off	got	white	scuff
40	neck	sick	as	frog	side
41	beef	thread	grew	bush	sled
42	few	day	knee	clown	smash
43	use	feel	fresh	cab	smooth
44	did	wood	tray	hurt	soap
45	hit	pig	cat	pass	stead
46	pond	crack	on	grade	taint
47	hot	dime	camp	blind	tap
48	own	wash	find	drop	thin
49	bead	and	yes	leave	tip
50	shop	look	loud	nuts	wean

Haskins' actual data (Percent Correct vs Sensation Level) are replotted in Figure 1 (adapted from Haskins, 1949, Figure 7). The data points as well as the best-fitting sigmoidal functions are plotted in this figure. The best-fitting curves are described as follows:

$$P(C)=\frac{a}{1+e^{-\frac{1}{b}(level-level_0)}}$$

where P(C) is the percentage of correct responses, a is the maximum possible P(C), $level_0$ is the level at the midpoint of the function, and b is the slope of the curve at the midpoint. The five curves in Figure 1 correlate with the individual data points nearly perfectly, ($r > +.99$).

Table 2.

Mean percent correct responses and standard deviations for PBK Lists 1-4 and PB-50 List 13. Values averaged across 22 adult listeners with normal hearing (adapted from Haskins, 1949, Table VI).

		PBK List 1		**PBK List 2**		**PBK List 3**		**PBK List 4**		**PB-50 List 13**	
Attenuation Level	Sensation Level	Mean	S.D.	Mean	S.D.	Mean	S.D	Mean	S.D	Mean	S.D.
70	-7.8	3.4	3.8	7.6	8.9	6.7	6.3	6.8	5.7	5.7	5.6
65	-2.8	16.3	10.3	24.2	10.8	16.8	8.9	19.5	9.1	17.5	8.2
60	2.2	33.8	10.6	44.1	11.9	36.6	9.5	39.3	12.5	33.4	8.2
50	12.2	69.9	8.8	82.2	11.2	71.5	11.2	73.2	9.0	68.2	7.8
35	27.2	98.1	3.1	98.4	1.7	97.1	2.8	97.7	2.4	95.7	1.9

Table 3.

Lexical neighborhood measures for PBK Lists 1-4 and PB-50 List 13.

	PBK List 1	**PBK List 2**	**PBK List 3**	**PBK List 4**	**PB-50 List 13**
Mean (Word Frequency)	558.6	954.4	689.1	612.7	295.6
Mean (Neighborhood Frequency)	294.6	261.4	231.5	272.5	196.9
Mean (Neighborhood Density)	17.0	16.4	17.5	17.7	17.3
Mean (Word Frequency/Neighborhood Frequency)	4.4	8.0	2.0	3.3	1.9
Mean (Word Frequency/Neighborhood Density)	35.3	94.1	42.5	41.3	13.0
Median (Word Frequency)	62.0	110.0	46.0	75.0	13.0
Median (Word Frequency/Neighborhood Frequency)	0.9	1.4	0.9	0.9	0.2
Median (Word Frequency/Neighborhood Density)	3.9	5.3	2.7	5.0	0.7

In examining the curves in Figure 1, one can see that performance on the PBK Test is clearly related to the presentation level of the stimulus. The slopes of the articulation functions between 20-80 percent correct response are approximately 4%/dB. At the ends of the functions (0-20%, 80-100%), performance changes little with a change in the level of presentation (approximately 1%/dB). Thus, for listeners with hearing losses, increasing the presentation level of the stimulus words by 5 dB could amount to as much as a 20% increase in performance, or very little increase in performance depending upon how much useable hearing the subject has and what region of the curve the subject is working in.

Insert Figure 1 about here

Haskins also examined test-retest reliability for her word lists at a single presentation level (12.2 dB SL). She found a large amount of improvement between the first and second repetitions of the test with the average score for List 2 increasing approximately 5.5%, and the average score for the remaining lists increasing approximately 8.8%. Significant improvement was seen for Lists 1 and 4 ($p < .05$), Lists 3 and 13 ($p < .01$), but not for List 2 ($p > .05$).

The results of the computational analysis of the lexical properties of the words on the PBK lists are displayed in Table 3. Here, we show both the mean and median word frequency, neighborhood frequency, and neighborhood density for the test items on the five word lists. In examining the mean word frequency, the reader will notice a very large difference in the mean frequency between List 2 (954.4) compared to the mean frequencies of the other three lists (295.6 - 689.1). Upon closer inspection, List 2 contains the word "and" which is one of the six most common words in the English language. The frequency of "and" is very high and skews the frequency distribution of List 2 a great deal. However, if we examine the median word frequency for the different lists, we still find that List 2 has a higher median frequency (110) than the other three lists (13 - 75). Thus, the words in List 2 do occur more frequently in the language than the words in the other lists, which may account, in part, for why List 2 was "more audible" than the other PBK lists.

Little difference exists in the mean neighborhood frequencies or densities between List 2 and the other test lists. The mean neighborhood frequency of List 2 (261.4) is within the range of mean neighborhood frequencies for the other three lists (196.9 - 294.6). The mean neighborhood density of List 2 (16.4 words) is slightly less than the mean neighborhood densities of the other lists (17.0 - 17.7 words). However, these differences are not very large. Thus, it appears that the mean neighborhood frequency and density of the words for the different PBK lists do not provide us much useful information about the reasons for the differences in audibility observed between the lists. However, these three measures (word frequency, neighborhood frequency, and neighborhood density) are absolute values for the words themselves without regard to context and the effects of other phonetically similar words. Put another way, these are "first-order" computational measures that do not take into account the relational properties of words to other phonetically similar words in the lexicon.

The next step in the computational analysis of the lexical properties of the PBK words was to generate ratios of the individual word frequencies to the neighborhood frequencies and densities of the individual words. These measures are "second-order" statistics that capture the relational properties of words to their lexical neighborhoods. Both mean and median values are listed in Table 3, but because of the effect of very high-frequency words (e.g., "and"), we will just examine the median values. Ratios were generated by dividing the frequency of the individual words by the neighborhood frequency of the

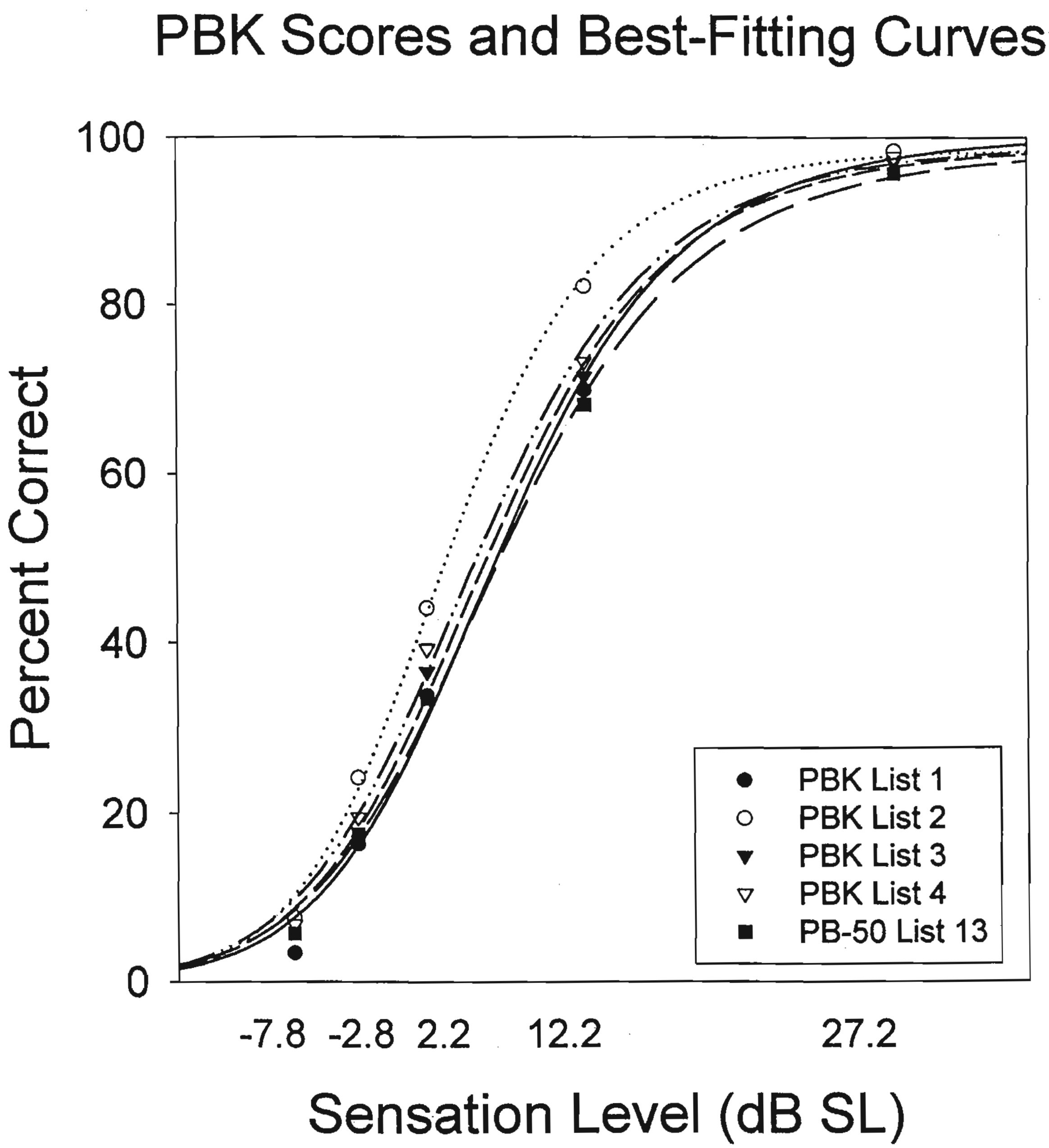

Figure 1. Percentage of correct responses vs Sensation Level for PBK Lists 1-4 and PB-50 List 13. Best-fitting curves are 3-parameter sigmoidal functions described in Equation 1. Adapted from Haskins (1949) Figure 7.

individual words. The median of these ratios for List 2 was approximately 1.4 while the median ratios for the three other PBK lists (0.2 - 0.9) were much smaller than the median ratio for List 2. This finding demonstrates that the words in List 2 have a higher frequency of occurrence than the other words in their lexical neighborhoods whereas the words in the remaining lists have a lower frequency of occurrence than the words in their lexical neighborhoods. Put another way, this analysis demonstrate that there is less lexical competition among phonetically similar words on PBK List 2 than the words on the other three PBK lists.

A second statistic was generated by dividing the frequency of the individual words by the neighborhood density of the individual words. The results of this analysis showed that the median ratio for List 2 (5.3) was larger than the median ratios for the other three PBK lists (0.7 - 5.0). As reported earlier, the median neighborhood densities for the different PBK lists were quite similar (range 16.4-17.7 words), however, this statistic provides further evidence that there is less lexical competition among phonetically similar words on PBK List 2 than the words on the other three PBK lists. Thus, words on List 2 should be less confusable than words on the other three lists. In short, these computational analyses of the similarity spaces or lexical neighborhoods of the words demonstrate that the words in PBK List 2 are inherently more distinctive and discriminable than the words on the three other PBK lists. If we assume that in open-set tests spoken words are recognized relationally in the context of other phonetically similar words, these computational analyses would predict that the words on PBK List 2 should be more acoustically distinctive and therefore more easily recognizable than the words on any of the other three lists. This is precisely the result that Haskins found in her thesis in 1949 almost 50 years ago.

Discussion

The results from the present analysis of the specific words used on the PBK Test give further support to NAM (Luce, 1986; Luce & Pisoni, in press; Luce, Pisoni, & Goldinger, 1990). NAM is a model of speech perception and spoken word recognition whose underlying hypothesis is that words are recognized relationally in the context of other phonetically similar words. This model attempts to account for the effects of lexical activation and competition on spoken word recognition. In particular, it is assumed that a sound pattern activates multiple lexical items in memory and this pattern of activation affects lexical discrimination and subsequent recognition performance (Luce & Pisoni, in press).

NAM predicts that when a spoken word is perceived, it activates a set of representations of similar sounding words in memory. This set of words is called a "lexical neighborhood." Once the neighborhood of the stimulus is activated, the word recognition system must then discriminate among the activated neighbors or candidates and decide which of the neighbors best matches the stimulus. The lexical discrimination and decision process is influenced by three factors: (1) the frequency of the target word; (2) the number of phonetically similar words (neighbors) activated in memory by the stimulus pattern; and (3) the frequencies of occurrence of the neighbors in the similarity neighborhood. Because of increased competition among activated items in memory, stimuli from densely populated lexical neighborhoods are more slowly and less accurately recognized than stimuli from sparsely populated neighborhoods. In addition, the presence of high frequency neighbors produces increased competition among words that are similar to the target word, thereby also reducing recognition performance.

According to NAM, identification performance can be predicted by the following rule (from Luce, 1986):

$$p(ID)=\frac{p(Stim)*Freq_s}{p(Stim)*Freq_s+\sum_{j=1}^{n}p(Neighbor_j)*Freq_j}$$

where p(ID) is the probability of correction identification, $p(Stim)*Freq_s$ is the frequency-weighted probability of identifying the stimulus word based on acoustic-phonetic information, and $p(Neighbor_j)*Freq_j$ is the frequency-weighted probability of identifying a neighbor. According to this rule, the model predicts that stimulus words with few low-frequency neighbors will be identified most accurately (because of the relatively small denominator in Equation 2, which indexes the degree of lexical competition among neighbors activated in memory). The model also predicts that stimulus words with many high-frequency neighbors will be identified least accurately (relatively large denominator in Equation 2).

The present analysis of the items on the PBK Test revealed that the median values of word frequency for the words in List 2 are considerably higher than the median values of word frequency for the words in the other lists. According to Pollack, Rubenstein and Decker (1959), listeners are more likely to correctly identify a high-frequency word than a low-frequency word at unfavorable S/N ratios when the message set is unknown. Thus, one would expect, and NAM predicts that on the average, at unfavorable S/N ratios, listeners would be more likely to correctly identify a word from List 2 than a word on the other lists. Moreover, scores with List 2 should be better than scores for the other lists at points along portions of the psychometric functions seen in Figure 1. At the end of the functions where the sensation levels of the stimuli are very low, the effect of the frequency of the stimulus word on the identification of that word is not very important, because there is little useful information in the signal at a presentation level at which some of the words are just becoming intelligible (very low S/N ratio). At the other end of the psychometric functions, the presentation level of the stimuli is high enough (high S/N ratio) so that all the words are easily intelligible, and word frequency again has little effect on identification.

The results of our analyses also revealed that the ratio of the frequency of the stimulus to the frequency of the neighborhood of the stimulus was higher for List 2 than the other lists. If the frequency of the stimulus is relatively great compared to the frequency of the neighborhood of that stimulus, then the numerator of Equation 2 is relatively large compared to the denominator, and the probability of correctly identifying the stimulus should high. Again, this effect is more important at intermediate S/N ratios where the stimulus is neither too easy nor too difficult to identify. The ratio of the frequency of the stimulus to the density of the neighborhood was also more favorable for the words on List 2 than the words on the other three PBK lists. If the density, or number of items in a lexical neighborhood is low in relation to the frequency of the stimulus, then the denominator in Equation 2 is again relatively small when compared to the numerator, and the probability of a listener correctly identifying the stimulus should be high. Thus, we find that both absolute and relative lexical neighborhood measures obtained in the present study lend further support to NAM.

The findings reported here are of both clinical and theoretical interest. A more practical question arises from the results -- are the lexical neighborhood differences observed here important for the target group the test was designed for and under the conditions the test is normally given? First, as was noted earlier, the PBK Test was constructed to be a test of word identification for children, and especially children with hearing losses. However, Haskins measured the "equivalence" of the PBK lists using adult subjects with normal hearing, not children with normal hearing or hearing loss. Nevertheless, the PBK Test has proven to be a very useful test for children, and it has probably been one of the most important measures of open-set word recognition for children with CIs (Carney et al., 1991; Fryauf-Bertschy et al., 1992; Fryauf-Bertschy et al., 1997; Gantz, Tyler, Tye-Murray et al., 1994; Gantz, Tyler, Woodworth et al., 1994; Kirk et al., 1995; Miyamoto et al., 1993; Miyamoto et al., 1994; Osberger et al., 1991; Pisoni et al., submitted; Staller, Beiter et al., 1991; Staller, Dowell et al., 1991; Waltzman et al., 1994; Waltzman et al., 1995). Second, the PBK Test is usually given by an audiologist through either monitored live voice or a tape-recorded version of the test at approximately 70 dB SPL. The most intense level Haskins presented to her subjects was approximately 27 dB above the average spondee thresholds for her group of subjects. If we assume that an average Speech-Reception Threshold (SRT) for spondaic words is approximately 20 dB SPL (ANSI S3.6, 1989; Young, Dudley, & Gunter, 1982) for a group of young adults with normal hearing, then Haskins presented the PBK lists to her subjects at approximately 47 dB SPL. We would expect performance to be perfect or nearly perfect for normal-hearing adults at a presentation level of 70 dB SPL, regardless of the PBK list tested, and that no differences between the lists would be apparent at this level.

For children with normal hearing, we would also expect nearly perfect performance at 70 dB SPL assuming that the children were able to perform the task and actually knew the words being tested. For children with hearing losses, their performance should be related to the level at which the word lists are presented. If the child has useable hearing and is able to understand speech well, and if the PBK words are presented at a high enough level, the child should perform well (nearly 100%) on the PBK Test (and performance should be nearly identical for the four different PBK lists). If the PBK lists cannot be tested at an appropriate level, or the child does not have useable hearing, then we would expect performance on the PBK Test to be low, and we might expect to see differences emerge between scores on PBK List 2 and the other PBK lists based on the lexical differences between the lists reported above.

The words chosen for the PBK Test were selected to be within the vocabulary of a child of based on norms for preschool children (The International Kindergarten Union, 1928). As stated earlier, Kirk et al. (1995) found that a large number of words on the PBK Test were not in a database containing words in the vocabulary of young children (Logan, 1992). In a more recent study, Kluck, Pisoni and Kirk (in preparation) obtained speech perception scores from 22 three-year-olds and 8 four-year-olds using recorded versions of the PBK Test and the LNT. The words were presented to the children at approximately 70 dB SPL in an auditory-only format. The four-year-old children performed at near-ceiling levels for both tests with performance on the PBK Test just slightly lower than performance on the LNT. Performance for the three-year-old children was slightly lower than performance for the four-year-old children on both tests. These results from the children with normal hearing provide a benchmark for children with impaired hearing and CIs. Clearly, three- and four-year-old children with normal hearing are able to imitate and repeat all the test items on both lists without any difficulty.

Kluck et al. (in preparation) addressed the issue of word familiarity in the different lists by asking the parents of the children whether or not their child was familiar with the word, using a seven-point scale. Although Kirk et al. (1995) found that a large number of words on the PBK Test were not in a corpus of words based on the vocabularies of young children (Logan, 1992), Kluck et al. found that all the words on both the PBK Test and the LNT were reported to be very familiar to the 4-year-old children, with the mean familiarity rating of the words from the PBK Test rated 6.58 (out of 7) and the words from the LNT rated 6.87 (out of 7). The words were slightly less familiar to the 3-year-old children (mean familiarity of PBK words - 5.95, mean familiarity of LNT words - 6.74 out of 7). Thus, it appears that the words on the PBK Test are less familiar to the three-year-olds than to the four-year-olds, but most of the words on the PBK Test are familiar to preschool-aged children.

Although the PBK Test may have certain shortcomings, it is certainly a difficult and probably the most important test of open-set word recognition for deaf children both with and without CIs. The tasks involved in the correct repetition of a stimulus in an open-set test of word recognition place a number of information processing demands on deaf children. The listener must search and retrieve words from lexical memory by encoding phonemic differences based on information present in the speech signal without the aid of any external context or retrieval cues and then discriminate and select a pattern from a large number of equivalence classes in memory. The listener must then use this information to form an articulatory plan or a motor program in order to produce a verbal or motor response to the stimulus. The various perceptual and cognitive operations needed for spoken word recognition require access to a variety of memory codes and neural representations of speech and spoken language at different levels of analysis. The speed and efficiency of these information-processing operations, particularly as they might be employed in tasks requiring transformation and mapping from perception to production, will depend to a large extent on having representations of words in memory and organizing these representations systematically in a lexicon that can be accessed efficiently to provide different sources of information about the words in the language.

We know that pediatric CI users recognize words relationally in accordance with the predictions of NAM. This was demonstrated by Kirk et al. (1995) using the LNT, which was designed especially to examine this question. It is possible that the lexical neighborhoods of children with CIs are fundamentally different from the lexical neighborhoods of children with normal hearing. We predict that because of poorer discrimination abilities, the neighborhoods of the CI users will be broader (i.e., contain more words) than the neighborhoods of children with normal hearing. For example, let us assume that a CI user has difficulty discriminate the place of articulation of a consonant. If the stimulus word presented to the listener is "pan", that listener would have difficulty discriminating the stimulus from words that vary from the stimulus word in the place of articulation of the initial consonant (e.g., "tan" and "can"). For a listener with normal hearing, the lexical neighborhood of the word "pan" includes the words "tan" and "can" as they vary from the stimulus by a single phoneme substitution. For the CI user who has difficulty discriminating the place of articulation of consonants, the lexical neighborhood of the stimulus "pan" might also contain words in the lexical neighborhoods of "tan", and "can". Thus, the structure of the lexical neighborhoods of pediatric CI users may be quite different from the structure of the lexical neighborhoods of a listener with normal hearing.

Although the lexical neighborhoods of a pediatric CI user may be broader than the lexical neighborhoods of a listener with normal hearing, the average size of the receptive vocabulary of the pediatric CI user or deaf child is considerably smaller than the average size of the receptive vocabulary of a group of children of the same age with normal hearing (Geers & Moog, 1994; Miyamoto et al., 1992). A smaller vocabulary should lead, therefore, to lexical neighborhoods that are, in general, less dense (contain fewer words) than lexical neighborhoods for listeners with normal hearing and normal vocabulary development.

At the present, we are conducting an analysis of the errors made by pediatric CI users with two and three years of implant use on open-set tests of word recognition (Meyer, Wright, Chin, & Pisoni, 1998). This analysis should give us a better understanding of the neighborhood structure of the children with CIs and provide insight into how these similarity neighborhoods change with increased implant use. As the ability of a pediatric CI user to make fine distinctions between the different acoustic-phonetic properties of words improves, the listener's lexical neighborhood structure for a particular word should become more refined and decrease in size. Preliminary analyses of the error data provide support for this prediction.

Summary and Conclusions

The PBK Test has been used for almost 50 years to assess the open-set speech perception skills of young children. Originally, Haskins generated four lists of words, but after testing the intelligibility of her lists at different S/N ratios with a group of normal-hearing adult listeners, she concluded that one of the word lists (PBK List 2) was easier (i.e., "more audible") than the other three lists and should not be used clinically. In the present study, we examined the lexical properties of all of the words in the four lists of the PBK Test to determine if any differences could be found in the similarity spaces of the words used on these lists in terms of measures of word frequency and lexical density. The results of our computational analysis demonstrate that the words on List 2 are of higher frequency and come from less dense neighborhoods than the words in the remaining three PBK lists.

The results of this analysis lend further support to the hypotheses set forth by Luce and his colleagues in describing NAM (Luce, 1986; Luce & Pisoni, in press; Luce, Pisoni & Goldinger, 1990) that words are recognized relationally in the context of other words. Word frequency and lexical density control the recognition process for words presented in isolation.

References

American National Standards Institute. (1989). Specifications for audiometers. (ANSI S3.6-1989). New York: ANSI.

Carney, A. E., Osberger, M. J., Miyamoto, R. T., Karasek, A., Dettman, D. L., & Johnson, D. L. (1991). Speech perception along a continuum: From hearing aids to cochlear implants. In Feigin, J. A. & Stelmachowicz, P. G. (Eds.) *Pediatric Amplification.* (pp. 93-113). Omaha, NE: Boystown National Research Hospital.

Dunn, L., & Dunn, L. (1981). Peabody picture vocabulary test - revised. Circle Pines, MN: American Guidance.

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope, 58*, 955-991.

Fryauf-Bertschy, H., Tyler, R. S., Kelsay, D. M., & Gantz, B. J. (1992). Performance over time of congenitally deaf and postlingually deafened children using a multichannel cochlear implant. *Journal of Speech and Hearing Research, 35,* 913-920.

Fryauf-Bertschy, H., Tyler, R. S., Kelsay, D. M.,Gantz, B. J., & Woodworth, G. G. (1997). Cochlear implant use by prelingually deafened children: The influences of age at implant and length of device use. *Journal of Speech, Language, and Hearing Research, 40,* 183-199.

Gantz, B. J., Tyler, R. S., Tye-Murray, N., Fryauf-Bertschy, H. (1994). Long term results of multichannel cochlear implants in congenitally deaf children. In Hochmair-Desoyer, I. J. & Hochmair, E. S. (Eds.) *Advances in Cochlear Implants* (pp. 528-533). Vienna, Austria: Manz.

Gantz, B. J., Tyler, R. S., Woodworth, G. G., Tye-Murray, N., & Fryauf-Bertschy, H. (1994). Results of multichannel cochlear implants in congenital and acquired prelingual deafness in children: Five-year follow-up. *American Journal of Otology, 12*, 1-7.

Geers, A. E., & Moog, J. S. (1994). Spoken language results: Vocabulary, syntax, and communication. *The Volta Review, 96,* 131-148.

Greenberg, J. H., & Jenkins, J. J. (1963). Studies in the psychological correlates of the sound system of American English. *Word, 20,* 157-177.

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

Kucera, F., & Francis, W. (1967). *Computational Analysis of Present-Day American English.* Providence, RI: Brown University Press.

Lane, H. (1995). Acquisition of speech perception ability in prelingually DEAF children with a multi-channel cochlear implant. Letter to the Editor. *American Journal of Otology, 16,* 393-399.

Luce, P. A., & Pisoni, D. B. (in press). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing.*

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* (pp. 122-147). Cambridge, MA: MIT Press.

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

Miyamoto, R. T., Osberger, M. J., Robbins, A. M., Myres, W. A., Kessler, K., & Pope, M. (1992). Longitudinal evaluation of communication skills of children with single- or multichannel cochlear implants. *American Journal of Otology, 13,* 215-222.

Miyamoto, R. T., Osberger, M. J., Robbins, A. M., Myres, W. A., & Kessler, K. (1993). Prelingually deafened children's performance with the Nucleus multichannel cochlear implant. *American Journal of Otology, 14*, 437-445.

Miyamoto, R. T., Osberger, M. J., Todd, S. L., Robbins, A. M., Stroer, M. A., Zimmerman-Phillips, S., & Carney, A. E. (1994). Variables affecting implant performance in children. *Laryngoscope, 104,* 1120-1124.

Osberger, M. J., Miyamoto, R. T., Zimmerman-Phillips, S., Kemink, 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 & Hearing, 12*, S66-S80.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4,* 75-95.

Pollack, I., Rubenstein, H., & Decker, L. (1959). Intelligibility of known and unknown message sets. *Journal of the Acoustical Society of America, 31*, 273-279.

Staller, S. J., Beiter, A. L., Brimacombe, J. A., Mecklenburg, D. J., & Arndt, P. (1991). Pediatric performance with the Nucleus 22-channel cochlear implant system. *American Journal of Otology, 12*, S126-S136.

Staller, S. J., Dowell, R. C., Beiter, A. L., Brimacombe, J. A. (1991). Perceptual abilities of children with the Nucleus 22-channel cochlear implant. *Ear & Hearing, 12*, S34-S47.

Svirsky, M. A. (1996). Speech production and language development in pediatric cochlear implant users. *Journal of the Acoustical Society of America, 99*, 2570.

Waltzman, S., Cohen, N., Gomolin, R., Ozdamar, S., Shapiro, W., & Hoffman, R. (1995). Effects of short-term deafness in young children implanted with the Nucleus cochlear prosthesis. *Annals of Otology, Rhinology, & Laryngology, 104*, 341-342.

Waltzman, S. B., Cohen, N. L., Gomolin, R. H., Shapiro, W. A., Ozdamar, S. R., & Hoffman, R. A. (1994). Long-term results of early cochlear implantation in congenitally and prelingually deafened children. *American Journal of Otology, 12*, 9-13.

Young, L. L., Dudley, B., & Gunter, M. B. (1982). Thresholds and psychometric functions of the individual spondaic words. *Journal of Speech and Hearing Research, 25*, 586-593.