Published in final edited form as:

Ear Hear. 2000 December ; 21(6): 578–589.

Modeling Spoken Word Recognition Performance by Pediatric Cochlear Implant Users using Feature Identification

Stefan A. Frisch and David B. Pisoni

Department of Linguistics (S.A.F.), University of Michigan, Ann Arbor, Michigan; and DeVault Otologic Research Laboratory (D.B.P.), Department of Otolaryngology, Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana. David B. Pisoni is also with the Speech Research Laboratory, Department of Psychology, Indiana University.

Abstract

Objective—Computational simulations were carried out to evaluate the appropriateness of several psycholinguistic theories of spoken word recognition for children who use cochlear implants. These models also investigate the interrelations of commonly used measures of closed-set and open-set tests of speech perception.

Design—A software simulation of phoneme recognition performance was developed that uses feature identification scores as input. Two simulations of lexical access were developed. In one, early phoneme decisions are used in a lexical search to find the best matching candidate. In the second, phoneme decisions are made only when lexical access occurs. Simulated phoneme and word identification performance was then applied to behavioral data from the Phonetically Balanced Kindergarten test and Lexical Neighborhood Test of open-set word recognition. Simulations of performance were evaluated for children with prelingual sensorineural hearing loss who use cochlear implants with the MPEAK or SPEAK coding strategies.

Results—Open-set word recognition performance can be successfully predicted using feature identification scores. In addition, we observed no qualitative differences in performance between children using MPEAK and SPEAK, suggesting that both groups of children process spoken words similarly despite differences in input. Word recognition ability was best predicted in the model in which phoneme decisions were delayed until lexical access.

Conclusions—Closed-set feature identification and open-set word recognition focus on different, but related, levels of language processing. Additional insight for clinical intervention may be achieved by collecting both types of data. The most successful model of performance is consistent with current psycholinguistic theories of spoken word recognition. Thus it appears that the cognitive process of spoken word recognition is fundamentally the same for pediatric cochlear implant users and children and adults with normal hearing.

Cochlear implants provide an opportunity for people with profound sensorineural hearing loss to achieve very high levels of speech perception performance (Skinner et al., 1994; Waltzman, Cohen, Gomolin, Shapiro, Ozdamar, & Hoffman, 1994). However, there is a great deal of variation between individuals, and many factors appear to influence the degree to which people who use cochlear implants can understand spoken language (Fryauf-Bertschy, Tyler, Kelsay, Gantz, & Woodworth, 1997). For children with prelingual profound

sensorineural hearing loss, cochlear implants provide an opportunity to acquire high proficiency in spoken language (Waltzman & Cohen, 2000).

One clinical goal of research on speech perception by pediatric cochlear implant users is to understand what type of sensory input is necessary for normal acquisition to proceed for most children who use implants, as it does among children without hearing impairment. Language acquisition by children who use a cochlear implant also raises several interesting psycholinguistic issues. It is unclear what measures and models are the best to use to assess the speech perception abilities of pediatric cochlear implant users. Further, very little research has investigated the impact of language acquisition with a cochlear implant on the cognitive abilities that are an integral part of speech perception and production (Pisoni, 1999).

The present study was carried out to try to predict open-set word recognition performance on the Phonetically Balanced Kindergarten (PBK) test (Haskins Reference Note 1) and Lexical Neighborhood Test (LNT) (Kirk, Pisoni, & Osberger 1995) by pediatric cochlear implant users based on performance in a closed-set phoneme identification task (see Table 1 for a glossary of acronyms that are frequently used in this paper). Three different computational models were used to explore the relative contribution of the different cognitive processes involved in spoken word recognition.

It is well known that adults with normal hearing recognize words by accessing a mental lexicon, and that performance is influenced by the usage frequency and density of words phonologically similar to the target word (Luce & Pisoni, 1998). Evidence that children with normal hearing and pediatric cochlear implant users have analogous lexical organization has also been found (Kirk et al. 1995). However, the role that phoneme recognition plays in spoken word recognition is an open issue. Some psycholinguistic theories use a phoneme-based lexical search, and others use phonetic or feature information to activate lexical candidates in memory (Forster, 1994). These theories vary in the extent to which phonemic information interacts with the perceptual space of words in the mental lexicon. Our three models of word recognition contrast in the role that phonemic information plays in the process of lexical access, in line with these theories.

The role of phonemic information in spoken word recognition also has clinical significance. Open-set word recognition performance is often scored by phonemes correct, rather than words correct. In addition, measures of speech perception performance that focus on phoneme identification are commonly used in evaluating children who use cochlear implants (e.g., Busby, Dettman, Altidis, Blamey, & Roberts, 1990). Although such measures provide important indicators of some speech perception abilities and may be necessary for developmental reasons, everyday language use requires successful word identification and lexical selection (Dowell & Cohen 1997). Thus, it is important to understand the relation between phoneme identification and spoken word recognition.

The present study focuses on children with profound sensorineural hearing loss who lost their hearing before the age of 2 and now use a cochlear implant. All children use the Nucleus 22-channel device with either the MPEAK coding strategy or the SPEAK coding strategy (Skinner et al., 1991, 1994; Staller, Beiter, & Brimacombe, 1994). These two groups of children are of interest for several reasons. First, they provide a wide range of performance levels over which to evaluate the relation between phoneme identification and open-set word recognition. Second, these children have acquired spoken language entirely through their cochlear implant. Their spoken word recognition performance therefore provides an important test of the generality of current hypotheses about spoken language processing in children and adults with normal hearing. It may be that no model of spoken

word recognition based on adults with normal hearing is appropriate for children who use implants. Third, although MPEAK and SPEAK users are quantitatively different in their language abilities, we wish to investigate whether these groups are qualitatively different in their performance (Parkinson, Parkinson, Tyler, Lowder, & Gantz, 1998; Skinner, et al., 1994). In other words, can the differences across groups be predicted from quantitative differences in their test performance with a single model, or are different models needed to account for the performance of each population?

Models

We developed a probabilistic model of phoneme recognition based on performance in feature identification obtained from the Minimal Pairs Test (Robbins, Renshaw, Miyamoto, Osberger, & Pope 1988). Phoneme confusions were predicted by assessing the contrastiveness of different types of linguistic features in the phoneme inventory of English. A probabilistic confusion matrix was created for each participant based on their performance on the feature identification task. This confusion matrix was then used as the input component to three models of word recognition. The first model, called the Phoneme Confusion Model or PCM, predicts that a word is recognized only on the basis of independently identifying the individual phonemes in the word. This model is based solely on phoneme confusion and has no lexical component.

The second model uses the output of PCM to perform a dictionary search and provides the best matching lexical item as an output response. This model is called SPAMR (Syllable Position Alignment for Matching and Retrieval). SPAMR uses PCM to make an early decision about the identity of phonemes, and then finds the best lexical match (cf. Taft & Forster, 1975).

The third model is a simplified computational implementation of the Neighborhood Activation Model (NAM; Luce, 1986; Luce & Pisoni, 1998). In this model, lexical items are activated directly by the input features, based on the confusion matrix derived from feature identification. The primary difference between SPAMR and NAM is that NAM delays decisions about phoneme identity until word recognition, so phoneme identification in NAM is interactive, not independent. The PCM, SPAMR, and NAM models provide a useful range of hypotheses about the relative importance of phonetic and lexical information and how they interact in the process of spoken word recognition. The remainder of this section discusses the implementation and assumptions of these models in detail.

Predicting Phoneme Confusions from Linguistic Features

The Minimal Pairs Test is a 2-alternative forced-choice word identification test containing words that contrast on a single phoneme (Robbins et al. 1988). A child is presented with one word on each trial and responds by selecting one of two alternative pictures. The set of minimally contrastive phonemes covers a variety of linguistic features. The dimensions of phonemic contrast in English can be grouped into six broad phonetic categories. Consonants contrast on place, manner, and voicing dimensions. Vowels contrast in vowel place, vowel height, and vowel manner dimensions. Five of these dimensions of contrast are represented in the Minimal Pairs Test. The phonemic contrasts used to represent these categories are shown in Table 2*. We assume that failure to differentiate contrasts for one dimension, such

*Note that some of the contrasts in the Minimal Pairs Test involve more than a single feature category. For example, the t/ʃ contrast is both a manner and (minor) place difference. The i/ɔ contrast is both a vowel place and vowel height difference. We assume the confounded features do not greatly affect the estimates of feature identification performance. The u/I contrast is both a vowel place and tense/lax difference. The Minimal Pairs Test does not directly investigate confusability along the tense/ lax dimension, which we include in our vowel manner dimension. For the simulations in this paper, vowel manner features are assumed to be discriminated at the average level of the vowel place and vowel height features.

as the place of articulation, for one consonant pair (e.g., p/k) indicates that discrimination for analogous pairs (e.g., b/g, k/t) is also difficult. Although some of the contrasts that have been grouped into a single category may have very different acoustic realizations (e.g., place in p/k versus f/*, manner in m/b versus f/p, voicing in t/d versus s/z), this study assumes for simplicity that all contrasts within a class are equivalent and that performance on any one contrast can be predicted from the average performance over the phoneme pairs that test that contrast.

To make predictions of phoneme confusions from the Minimal Pairs Test, it is necessary to decide explicitly whether a failure to identify some contrast will result in confusions between a particular pair of phonemes. To answer this question, we turn to a set-theoretic model of the phonemic inventory called structured specification (Broe, Reference Note 2). Using structured specification, we can directly predict phoneme confusions if some contrasts cannot be discriminated (see Frisch & Pisoni, 1997, for a more extensive discussion).

For example, consider a five vowel inventory /i, e, a, o, u/. All the phonemes in this inventory can be distinctly described by the features given in (1). The phoneme /e/ is identified by the combination of the vowel place feature [1front] and the vowel height feature [1mid]. Now suppose that no vowel height features are discriminable. This state of affairs can be simulated by removing the vowel height features from the inventory. With no vowel height features there are two vowel classes: /i, e/ and /a, o, u/, defined by the remaining vowel place features. We can interpret this to mean /i/ and /e/ are confusable, and /a/, /o/, and /u/ are confusable.

	vowel place		vowel height		
	[front]	[back]	[high]	[mid]	[low]
i	+		+		
e	+			+	
a		+			+
o		+		+	
u		+	+		

(1)

For simplicity, it is assumed that when several phonemes are confusable, each is equally likely to be identified as the perceived phoneme. A confusion matrix for the five vowel inventory when vowel height features are removed is shown in (2). The intended phoneme (stimulus) is given in the left column, the perceived phoneme (response) is given across the top. A phoneme is always assumed to be confusable with itself. In other words, there is a possibility that the correct phoneme will be identified even if a contrast cannot be perceived.

	i	e	a	o	u
i	0.46	0.10	0.06	0.06	0.30
e	0.10	0.46	0.06	0.30	0.06
a	0.06	0.06	0.69	0.09	0.09
o	0.06	0.30	0.09	0.45	0.09
u	0.30	0.06	0.09	0.09	0.45

(2)

Confusions for more complicated phoneme inventories, like the full set of vowels and consonants in English, can be predicted from features in exactly the same way. The

phoneme inventory and feature sets used for the computational models in this paper are given in the Appendix. These features and the model of confusability given here have also been successfully applied to model errors in speech production and to model phonological constraints based on phoneme confusability (see Frisch, Reference Note 3).

Predicting Phoneme Identification

The Minimal Pairs Test is a 2-alternative word identification task. The first step in using performance on this task to predict open-set word recognition performance is to transform the 2-alternative word identification scores into estimates of open-set feature identification performance. There are no studies we know of that have evaluated the relation between closed-set word recognition and open-set feature identification in this way. We assume that closed-set and open-set feature identification are equivalent processes, with the advantage of the restricted choices in a closed-set task factored out (Black, 1957; see Sommers, Kirk, & Pisoni, 1997, for related discussion).

The Minimal Pairs Test provides identification scores for individual feature contrasts. The simplest assumption for identification of several contrasts is that the identification of each contrast is independent. Under this assumption, the probability of simultaneously identifying multiple contrasts is the product of the probabilities of identifying each contrast individually. The probabilities for each possible outcome can then be used as weights for combining confusion matrices for each contrast or combination of contrasts. The result is a probabilistic confusion matrix that takes into account the statistical reliability of feature identification for individual features and their combinations. A sample probabilistic confusion matrix using the five vowel inventory and two dimensions of contrast will be worked out in detail to illustrate the computational procedure.

Suppose that 2-alternative closed-set identification in the five vowel inventory is 60% for vowel place, and 80% for vowel height. With the advantage of a 2-alternative choice factored out, open-set identification is estimated to be $(60-50)/(100-50) = 20\%$ for vowel place, and $(80-50)/(100-50) = 60\%$ for vowel height (Black, 1957). Based on these values, the estimated probability distribution for identifying contrasts is given in (3).

<u>Contrasts maintained</u>	<u>Estimated probability</u>
None	$(1-0.2) \cdot (1-0.6) = 0.32$
Vowel Place	$(0.2) \cdot (1-0.6) = 0.08$
Vowel Height	$(1-0.2) \cdot (0.6) = 0.48$
All	$(0.2) \cdot (0.6) = 0.12$

(3)

Each probability in the distribution in (3) is associated with an appropriate confusion matrix. To determine the cumulative matrix, each cell in each matrix is multiplied by the appropriate probability (weight), and corresponding cells are then summed across matrices. For example, given the confusion matrix in (2) and appropriate matrices for the identification of no contrasts, vowel height contrasts, and all contrasts, the probability of correct perception of /i/ would be $0.32 \times 0.2 + 0.08 \times 0.5 + 0.48 \times 0.5 + 0.12 \times 1 = 0.46$. The cumulative confusion matrix for this example is given in (4). This matrix reflects the initial probabilities used in the example. The example listener was better at detecting height differences than place differences, and the matrix predicts more place confusions than height confusions.

	i	e	a	o	u
i	0.5	0.5	0	0	0
e	0.5	0.5	0	0	0
a	0	0	0.33	0.33	0.33
o	0	0	0.33	0.33	0.33
u	0	0	0.33	0.33	0.33

(4)

Confusion matrices for the inventories of consonants and vowels in English can be generated following the exact same algorithm. For the consonants and for the vowels, eight probability weighted confusion matrices were added together based on the estimated probabilities of feature detection on the three dimensions of contrast for each phoneme type.

Modeling Spoken Word Recognition

Given a model of phoneme confusion for the open-set task, the simplest computational model of spoken word recognition applies the PCM on each phoneme in the word and treats the result as the recognized word. This PCM does not employ any lexical knowledge because the perceptual result is not matched to an internalized representation of words in the mental lexicon. This model of word recognition produces percepts such as those in (5) for the target word *please*. This simple model is certainly unrealistic as a model of spoken word recognition. Boothroyd and Nitttrouer (1988) concluded that such a model is appropriate for non-word recognition. We include this model as a benchmark in our simulations to gain insight into the performance of two more realistic models.

$$\begin{array}{ccccc}
 \text{fniz} & \text{pvi} & \text{pzo}^1\text{z} & \text{pli}\delta & \text{mlen} \\
 \text{wliz} & \text{rliz} & \text{pli} & \int\text{liz} & \text{dmiz} \\
 \text{pniz} & \text{tlez} & \text{bli} & \text{pdis} & \int\text{laz}
 \end{array} \quad (5)$$

For the two models that include a lexical access component, the phoneme confusion matrices still provide the core information on which confusions are based. In the SPAMR model, the output of PCM is used to conduct a lexical search, and the optimal match is chosen. The best match in SPAMR is determined by the number of shared phonemes between words aligned by onset, nucleus, and rime units. An online version of the Webster's Pocket Dictionary was used as the database for lexical searches (Nusbaum, Pisoni, & Davis, 1984). For example, for the target *please*, the PCM output [pdis] matches with *peace* except for one phoneme in the onset. No other word is as similar, so *peace* is chosen by SPAMR as the response for this percept. For simplicity, ties are categorically broken in favor of the more frequent word.

In the NAM, words in the lexicon are activated directly by phonetic information in the input, without an independent phonemic representation (Luce 1986; Luce & Pisoni 1998). The response is determined by a decision rule that combines word frequency and phonetic similarity. In this paper, we use a simplified version of NAM where words are activated by perceived features, and frequency effects are ignored. Given the phoneme confusion matrices for an individual listener, the likelihood of perceiving a target stimulus as a particular lexical item can be computed by determining the probability of confusion for each segment, and multiplying together these probabilities. For example, the likelihood of responding *fresh* for the stimulus *please* is computed as in (6). The response for an input in

the simplified NAM is chosen probabilistically, based on the likelihood of each word occurring as a confusion for the target.

$$p([frɛf] | [plɪz]) = p(f|p) \cdot p(r|l) \cdot p(ɛ|i) \cdot p(f|z) \quad (6)$$

For simplicity, confusions in NAM are also assumed to have the same number of consonants and vowels in the same syllable positions as the target. Note that confusions for phonemes are not completely independent in NAM, as some combinations of phonemes are not found in the lexicon. These combinations are not a part of the resulting distribution of possible phoneme confusions, so the actual distribution of phoneme confusions can be influenced by phonotactic constraints (Luce, 1986).

Simulation and Results

The procedure exemplified above for the recognition of *please* was applied to each of the words in the PBK and LNT tests of open-set spoken word recognition based on performance in feature identification using data from the Minimal Pairs Test for individual pediatric cochlear implant users.

Participants

The pediatric cochlear implant users examined in this study were selected from the population of children followed at the Indiana University School of Medicine. Forty-eight children who used the Nucleus 22-channel processor and either the MPEAK or SPEAK processing strategy were candidates for the present study. Data were examined for all such children who had been given the Minimal Pairs Test and at least one open-set word recognition test in the interval from 1.5 to 2 yr after they were fit with their implant. In cases where a child was tested twice during this interval, an average of the two test scores was used. Performance was only modeled when there was behavioral data available for comparison.

Children in both groups had prelingual sensori-neural hearing loss, but the groups differed on some demographic variables. These groups provide a range of performance on the Minimal Pairs Test, and PBK and LNT word recognition tests. Broad demographic variables, aggregate performance on the behavioral tasks, and the number of participants simulated in each group are given in Table 3.

The MPEAK and SPEAK groups are well-matched for their chronological ages. As was mentioned above, the data were selected from 1.5 to 2 yr postimplant, so the children are also well-matched for age at implantation. There are some differences in age of onset of sensorineural hearing loss, but all children had prelingual hearing loss. There is a relatively larger proportion of MPEAK users who participate in total communication education programs compared with the SPEAK users, who are fairly evenly balanced between total communication and oral programs. The MPEAK users also have a statistically higher unaided hearing threshold than the SPEAK users; however, all of the children have profound sensorineural hearing loss. Although it is not a focus of this study, these variables may play a role in explaining the differences between individuals in their performance on the feature identification and word recognition tests (van Dijk, van Olphen, Langereis, Mens, Brokx, & Smoorenburg 1999; Fryauf-Bertschy et al., 1997; Ouellet & Cohen 1999). However, as we show below, there appear to be no qualitative differences in performance between the two groups, and so the individual differences do not appear to affect the basic psycholinguistic processes involved in spoken word recognition.

Materials

The simulations modeled word recognition performance for the 150 words on the PBK and for the 50 “easy” and 50 “hard” words on the LNT. In the LNT test, easy and hard word lists differ in the lexical characteristics of the items (Kirk et al. 1995). Hard words are more difficult to recognize than easy words because the hard words are lower in usage frequency and confusable with a greater number of other words in comparison with the easy words. The PBK, LNT-Easy, and LNT-Hard lists given to the children were 25 word sublists from each test. The PBK and LNT tests were administered to the children using live voice presentation in an auditory only mode during a routine clinical visit in which many other measures of speech and language ability were also gathered.

Methods

As a preliminary test of the feasibility of using feature identification performance to predict open-set word recognition, correlations between performance on the Minimal Pairs Test and the LNT and PBK tests were examined. Performance on the Minimal Pairs Test was then used to generate a confusion matrix for vowels and consonants for each participant. Word recognition was simulated for 50 repetitions of each test using each of the three models for each participant.

Results

Correlations between the average performance across different segmental contrasts on the Minimal Pairs Test and performance on the LNT and PBK tests are shown in Table 4. The presence of strong, positive, and significant correlations suggests that an association between closed-set feature identification and open-set word recognition is present in these data sets. Thus, we can conclude that the models have the opportunity to capture some of the variation between individuals within each group.

Mean actual and predicted performance on the LNT-Easy word list is shown in Figure 1. Performance scored as percent correct words is shown in the top panel and performance scored as percent correct phonemes is shown in the bottom panel. In each panel, actual and model performance for the MPEAK users is shown on the left and actual and model performance for the SPEAK users is shown on the right. The model fit is evaluated statistically using paired *t*-tests for actual and predicted performance (denoted t_M for MPEAK users and t_S for SPEAK users throughout). For percent words correct, the NAM model provides a very good prediction of actual performance ($t_M(17) = 0.6$, $t_S(15) = 0.5$). The PCM and SPAMR models both significantly under-predict actual words correct (for PCM $t_M(17) = 5.5$, $p < 0.01$; $t_S(15) = 9.5$, $p < 0.01$; for SPAMR $t_M(17) = 5.2$, $p < 0.01$; $t_S(15) = 6.2$, $p < 0.01$). For percent phonemes correct, NAM over-predicts performance ($t_M(17) = 2.8$, $p < 0.05$; $t_S(15) = 3.9$, $p < 0.05$) and PCM and SPAMR under-predict performance (for PCM $t_M(17) = 2.6$, $p < 0.01$; $t_S(15) = 3.7$, $p < 0.01$; for SPAMR $t_M(17) = 5.8$, $p < 0.01$; $t_S(15) = 2.1$, $p < 0.05$).

Mean actual and predicted performance for the two groups on the LNT-Hard word list, scored as percent correct words (top panel) and phonemes (bottom panel), is shown in Figure 2. For words correct, as in the case of the LNT-Easy list, NAM provides an accurate prediction of actual performance ($t_M(14) = 0.8$; $t_S(12) = 0.01$). The PCM and SPAMR models again under-predict words correct ($t_M(14) = 3.1$, $p < 0.01$; $t_S(12) = 4.6$, $p < 0.01$ and $t_M(14) = 2.9$, $p < 0.05$; $t_S(12) = 4.2$, $p < 0.01$, respectively). For phonemes correct, NAM over-predicts performance for both groups ($t_M(14) = 3.7$, $p < 0.01$; $t_S(12) = 2.3$, $p < 0.05$). The PCM and SPAMR models provide a good prediction of phonemes correct for the MPEAK users ($t_M(12) = 1.0$, $t_M(12) = 0.8$, respectively). However, the models under-predict

phonemes correct for the SPEAK users ($t_5(14) = 2.5, p < 0.05$; $t_5(14) = 2.6, p < 0.05$, respectively).

Mean actual and predicted performance for each group on the PBK word lists, scored in percent correct words (top panel) and phonemes (bottom panel), is shown in Figure 3. For words correct, actual performance appears to be in between the higher predictions of NAM and the lower predictions of PCM and SPAMR. NAM significantly over-predicts words correct ($t_M(26) = 5.3, p < 0.01$; $t_5(18) = 4.3, p < 0.01$). The PCM model significantly under-predicts words correct, although it comes somewhat closer than NAM ($t_M(26) = 3.4, p < 0.01$; $t_5(18) = 2.1, p < 0.05$). The SPAMR model appears to make the best prediction of words correct. Although it under-predicts performance for the MPEAK users ($t_M(26) = 2.1, p < 0.05$), there is no significant difference between actual and predicted words correct for SPEAK users ($t_5(18) = 1.0$). For phonemes correct, NAM significantly over-predicts performance ($t_M(26) = 6.3, p < 0.01$; $t_5(18) = 5.0, p < 0.01$). The PCM and SPAMR models provide accurate predictions of actual performance ($t_M(26) = 0.2, t_5(18) = 1.3$ and $t_M(26) = 0.7, t_5(18) = 1.2$, respectively).

To confirm that the models also captured some of the variation between individuals, correlations between model predictions and performance data were computed. The correlations for model predictions were slightly higher ($r = +0.60 - 0.82$) and at the same significance level as the correlations for the raw Minimal Pairs Test scores in Table 4, suggesting that some individual variation in performance was indeed captured in the predicted scores.

Discussion

Despite the somewhat complex results, it is clear that there were no qualitative differences between the children who use MPEAK and the children who use SPEAK. Although the SPEAK users generally performed better than MPEAK users, as has been found elsewhere (Parkinson et al. 1998; Skinner et al. 1994), the difference in performance appears to be quantitative, not qualitative. The similarities and differences between actual performance and the predictions of the psycholinguistic models were fundamentally the same in almost all cases. The one exception was performance on the LNT-Hard word list scored by phonemes correct, where the PCM and SPAMR models predicted performance accurately for the MPEAK users only. We believe this outcome was coincidental, as the broader pattern of results indicates that NAM is the most appropriate model of spoken word recognition performance for both groups. The PCM and SPAMR models appear to be better suited as models of processing nonwords and unfamiliar words, as we discuss below.

Although the results were somewhat mixed across models and word lists, several generalizations emerged. The first and most straightforward pattern in the results is that predictions of performance by PCM and SPAMR were similar, and clearly different from the predictions of NAM. PCM and SPAMR make early, independent phoneme decisions based on feature input. PCM treats word recognition as equivalent to phoneme recognition, with no lexical access at all. SPAMR performs a lexical search after phoneme decisions are made. But this lexical search did not do much to change the predicted performance in comparison with PCM. Predictions of words correct were slightly higher for SPAMR than PCM, whereas predictions of phonemes correct were identical for the two models.

The second pattern that emerged from these analyses is that the performance of the models in predicting words correct was different for the LNT and PBK word lists. For both LNT lists, NAM gave accurate predictions. In contrast, none of the three models fit the PBK data especially well. The SPAMR model came the closest. Given its good performance in

predicting words correct on the LNT, we tentatively conclude that NAM provides the best model of the process of spoken word recognition. This suggests that phoneme decisions are delayed during word recognition. In other words, phoneme identification in word recognition occurs only as a part of a more global process of lexical access. This conclusion does not imply that phoneme recognition does not take place at all, because it is also possible to identify nonwords and recognize the phonetic content of a novel pattern. However, the pattern of results does imply that phoneme decisions are made interactively with word decisions during word recognition (Luce & Pisoni, 1998).

There are several important differences in the words used on the LNT and PBK lists that may account for the divergent results between lists. The test words on the LNT were selected from the CHILDES database of child speech and child-directed adult speech (MacWhinney, 1996). The words chosen for use on this test are familiar to young children with normal hearing. The PBK was created in the late 1940s, well before such a database existed (Meyer & Pisoni, 1999). Recent findings suggests that many of the words on this test may be unfamiliar to young children with sensorineural hearing loss (Kirk et al. 1995; Seghal, Kirk, & Hay-McCutcheon, in press). For some words on the PBK, the children may not have knowledge of a corresponding lexical item or a representation of the word in their lexicons. Thus, children given the PBK test may be using a mixture of word recognition and nonword naming abilities, depending on the particular stimulus items. If we consider PCM to be an appropriate model of non-word naming and NAM to be an appropriate model of lexical access and word recognition, we would predict that word recognition performance on the PBK test would be somewhere in between PCM and NAM. This is exactly what we found.

The third pattern that we observed was that the results are completely different when phonemes correct are analyzed instead of words correct. In general, actual performance as measured by phonemes correct was in between the lower predictions made by PCM and SPAMR, and higher predictions made by NAM. NAM always over-predicted performance, whereas PCM and SPAMR were successful in predicting phonemes correct on the PBK and for MPEAK users on the LNT-Hard word list. Given that the PBK test may involve naming and recognition processes more like non-word repetition than familiar word recognition, the better predictions of PCM for the PBK test can be straightforwardly explained. The fact that PCM does not employ lexical knowledge makes it a more appropriate model in the case of non-word repetition.

This leads us to the question of how it is that NAM can provide an accurate prediction of words correct but over-predict phonemes correct. The answer may be that when NAM chooses an incorrect response, it generally chooses one that is more similar to the target word than the erroneous responses given by the children. There are two potential explanations for how this could happen. It could be that the NAM simulation uses feature information more effectively than the children do. There is some evidence from studies with adults who use cochlear implants that their performance in speech perception and word recognition tasks is suboptimal given their psychophysical abilities (Meyer, Svirsky, Frisch, Kaiser, Pisoni, & Miyamoto, 1999). Using models similar to those in the present study, Meyer et al. found that the actual performance of adult cochlear implant users was worse than their predicted performance given model inputs based on psychophysical performance. However, Meyer et al. found that the model over-predicted performance scored by words correct and by phonemes correct owing to an over-prediction of abilities overall. This was not found in the present study.

Alternately, it may be that the use of a dictionary as the database for lexical access provided more near misses to the target words than children actually know. A near miss would result

in a higher percent correct phonemes score, but not affect the words correct score. For example, the NAM simulations made responses such as *plead*, *plebe*, and *pleat* to the target *please*. It is very unlikely that these low-frequency words are familiar to young children. We investigated the possibility that lexicon size influences predicted phonemes correct in NAM by eliminating approximately half of the lexicon and conducting new simulations. Words were eliminated from the lexicon according to their usage frequency. With the reduced lexicon, the predicted level of phonemes correct in NAM generally was reduced, whereas words correct was not significantly affected (although it did increase slightly). Thus, it is very likely that the predictions of NAM for phonemes correct would be more accurate if a better approximation to the smaller lexicons of young children were available.

The difference in the predictions of NAM depending on whether a larger or smaller lexicon is used leads to an interesting implication for clinical evaluation of spoken word recognition ability. The NAM model predicts that performance scored by phonemes correct is sensitive to lexicon size. If NAM is an appropriate model of word recognition for children who use cochlear implants, improvement over time in word recognition scored by phonemes correct could be the result of either better perceptual skills or greater lexical knowledge, as increased lexical knowledge will provide more near misses to the target words. Predicted performance scored by words correct was not significantly influenced by lexicon size, assuming of course that all the words on the test are in the smaller lexicon. On a test like the LNT where we can be reasonably sure that the test words are familiar to young children who use cochlear implants, scoring word recognition performance by words correct may better reflect genuine improvement in perceptual abilities.

Conclusions

Based on our computational simulations, we have found that children who use MPEAK or SPEAK go about the task of recognizing spoken words in fundamentally the same way. Further, by comparing observed performance with three models of spoken word recognition, we have some preliminary evidence that the process of spoken word recognition by pediatric cochlear implant users is much the same as in children and adults with normal hearing. Pediatric cochlear implant users are sensitive to the phonetic similarity of spoken words represented in an internalized lexicon, as predicted by current psycholinguistic theories like NAM. The PCM has no concept of a lexicon, and was thus unable to compensate for incorrect phonemic information that resulted in nonexistent words. Adding a later process of lexical matching in SPAMR did not do much to improve predicted performance. This finding suggests that phoneme recognition may be best viewed as a secondary result of the primary process of spoken word recognition.

The present findings have some broader implications for clinical work with children who are acquiring language using a cochlear implant. Given that the performance of children who use MPEAK and children who use SPEAK can be captured by a single model, it appears that the amount of phonetic information transmitted by these processing strategies is sufficient for children who use MPEAK and SPEAK to develop spoken word recognition processes in much the same way as children and adults with normal hearing.

The differences in the predictions of the three models highlight the important contribution of cognitive processes to speech perception and spoken word recognition. This approach contrasts with much research that focuses only on differences in phoneme recognition (e.g., Boothroyd, 1997; Fu & Shannon, 2000; Rubenstein & Miller, 1999; Svirsky, 2000). Successful implant users are able not only to discriminate distinct speech sounds, but more importantly, they are able to isolate, discriminate, select, and identify words from a multi-dimensional lexical similarity space. Although phoneme perception and word recognition

are closely related processes, they are not equivalent. Examining both sets of abilities and their interrelations may provide important new insights into the linguistic development of children who use cochlear implants and the large individual differences in speech and language measures that have been reported in the literature over the years.

Acknowledgments

We thank Steve Chin, Ph.D., Paul Luce, Ph.D., and four anonymous reviewers for helpful comments on an earlier version of this paper. Special thanks to Ted Meyer, M.D., Ph.D., for extensive discussion, comments, and criticism of this work. This research was supported in part by NIH Training Grant DC00012, R01-DC00111, and R01-DC00043 to Indiana University and in part by the *Language Learning* Visiting Research Assistant Professorship at the University of Michigan.

Appendix

The computational analysis in this paper uses the phoneme inventory in the online version of Webster's pocket dictionary (Nusbaum et al., 1984), known as the Hoosier Mental Lexicon or HML. Phonemes are indicated here with both IPA transcription and the equivalent HML character. Features used for the consonants and vowels, along with the broad categories of contrast by which they were grouped, are given in (A1) and (A2).

Features used to represent the English consonant inventory:

place:

IPA	p	b	f	v	m	t	d	θ	ð	s	z	ʃ	ʒ	tʃ	dʒ	k	g	ŋ	l	r	n	w	j	h
HML	p	b	f	v	m	t	d	T	D	s	z	S	Z	C	J	k	g	G	l	r	n	w	y	h
labial	+	+	+	+	+															+		+		
corona						+	+	+	+	+	+	+	+	+	+				+		+		+	
dorsal																+	+	+						
bilabial	+	+			+																		+	
dental			+	+				+	+															
alveolar						+	+			+	+								+	+	+			
palatal												+	+	+	+									+
velar																+	+	+						

manner:

IPA	p	b	f	v	m	t	d	θ	ð	s	z	ʃ	ʒ	tʃ	dʒ	k	g	ŋ	l	r	n	w	j	h
HML	p	b	f	v	m	t	d	T	D	s	z	S	Z	C	J	k	g	G	l	r	n	w	y	h
obstruent	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+							+
sonorant					+														+	+	+	+	+	+
stop	+	+			+	+	+							+	+	+	+	+						
continuant			+	+				+	+	+	+	+	+											+
glide																							+	+
consonantal					+														+	+	+	+		
oral	+	+				+	+									+	+							
nasal					+													+			+			
affricate														+	+									
strident										+	+	+	+	+	+									
distributed			+	+				+	+															+

[illegible]

Features used to represent the English vowel inventory:

<u>vowel place:</u>		IPA	i	ɪ	e	ɛ	æ	a	ʌ	u	ʊ	o	ɔ	oʰ	aʰ	ɔʊ	ə	ə	ɹ	ʒ	!	ŋ	ɱ	
	HML	i	I	e	E	@	a	^	u	U	o	c	O	Y	W	x		X	R	L	N	M		
	front	+	+	+	+	+	+							+	+			+				+		
	mid								+								+			+	+			
	back									+	+	+	+			+								
	labial									+	+	+	+	+		+							+	
<u>vowel height:</u>		IPA	i	ɪ	e	ɛ	æ	a	ʌ	u	ʊ	o	ɔ	oʰ	aʰ	ɔʊ	ə	ə	ɹ	ʒ	!	ŋ	ɱ	
	HML	i	I	e	E	@	a	^	u	U	o	c	O	Y	W	x		X	R	L	N	M		
	high	+	+							+	+			+	+	+			+		+			
	mid-high			+								+		+			+	+		+				
	mid-low				+			+				+				+								
	low					+	+								+									
<u>vowel manner:</u>		IPA	i	ɪ	e	ɛ	æ	a	ʌ	u	ʊ	o	ɔ	oʰ	aʰ	ɔʊ	ə	ə	ɹ	ʒ	!	ŋ	ɱ	
	HML	i	I	e	E	@	a	^	u	U	o	c	O	Y	W	x		X	R	L	N	M		
	static	+		+	+	+	+			+	+	+	+											
	dynamic													+	+	+								
	tense	+		+			+	+	+		+	+	+	+	+	+								
	lax		+		+	+					+													
	stressed	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+								
	unstressed																+	+	+					
	consonantal																				+	+	+	+

References

- Black JW. Multiple-choice intelligibility tests. *Journal of Speech and Hearing Disorders*. 1957; 22:213–235. [PubMed: 13429664]
- Boothroyd A. Auditory capacity of hearing-impaired children using hearing aids and cochlear implants: Issues of efficacy and assessment. *Scandinavian Audiology Supplement*. 1997; 26(Suppl. 46):17–25. [PubMed: 9309834]
- Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America*. 1988; 84:101–114. [PubMed: 3411038]

- Busby, PA.; Dettman, SJ.; Altidis, PM.; Blamey, PJ.; Roberts, SA. Assessment of communication skills in implanted deaf children.. In: Clark, GM.; Tong, YC.; Patrick, JF., editors. Cochlear prostheses. Churchill Livingstone; Edinburgh: 1990.
- van Dijk JE, van Olphen AF, Langereis MC, Mens LHM, Brokx JPL, Smoorenburg GF. Predictors of cochlear implant performance. *Audiology*. 1999; 38:109–116. [PubMed: 10206520]
- Dowell, RC.; Cowan, RS. Evaluation of benefit: Infants and children.. In: Clark, GM.; Cowan, RS.; Dowell, RC., editors. Cochlear implantation for infants and children: Advances. Singular Publishing Group; San Diego: 1997. p. 205-222.
- Forster KI. Computational modeling and elementary process analysis in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*. 1994; 20:1292–1310. [PubMed: 7844513]
- Frisch, SA.; Pisoni, DB. Predicting open-set spoken word recognition performance from feature discrimination scores in pediatric cochlear implant users: A preliminary analysis.. In: Pisoni, DB., editor. Research on Spoken Language Processing Progress Report No. 21. Indiana University; Bloomington, IN: 1997. p. 261-287.
- Fryauf-Bertschy H, Tyler RS, Kelsay DM,R, Gantz BJ, Woodworth GG. 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*. 1997; 40:183–199.
- Fu Q-J, Shannon R. Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners. *Journal of the Acoustical Society of America*. 2000; 107:589–597. [PubMed: 10641667]
- Kirk KI, Pisoni DB, Osberger M. Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing*. 1995; 16:470–481. [PubMed: 8654902]
- Luce, PA. Research on Spoken Language Processing Technical Report No. 6. Indiana University; Bloomington, IN: 1986. Neighborhoods of words in the mental lexicon..
- Luce PA, Pisoni DB. Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing*. 1998; 19:1–36. [PubMed: 9504270]
- MacWhinney B. The CHILDES system. *American Journal of Speech Language Pathology*. 1996; 5:5–14.
- Meyer TA, Pisoni DB. Some computational analyses of the PBK test: Effects of frequency and lexical density on spoken word recognition. *Ear and Hearing*. 1999; 20:363–371. [PubMed: 10466571]
- Meyer TA, Svirsky MA, Frisch SA, Kaiser AR, Pisoni DB, Miyamoto RT. Modeling closed-set phoneme and open-set word recognition by multi-channel cochlear implant users. *Journal of the Acoustical Society of America*. 1999; 106:2177.
- Nusbaum, HC.; Pisoni, DB.; Davis, CK. Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words.. In: Pisoni, DB., editor. Research on Spoken Language Processing Progress Report No. 10. Indiana University; Bloomington, IN: 1984. p. 357-376.
- Ouellet C, Cohen H. Speech and language development following cochlear implantation. *Journal of Neurolinguistics*. 1999; 12:271–288.
- Parkinson AJ, Parkinson WS, Tyler RS, Lowder MW, Gantz BJ. Speech perception performance in experienced cochlear-implant patients receiving the SPEAK processing strategy in the Nucleus Spectra-22 cochlear implant. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1073–1087.
- Pisoni, DB. Individual differences in effectiveness of cochlear implants in prelingually deaf children: Some new process measures of performance.. In: Pisoni, DB., editor. Research on Spoken Language Processing Progress Report No. 23. Indiana University; Bloomington, IN: 1999. p. 3-49.
- Robbins, AM.; Renshaw, JJ.; Miyamoto, RT.; Osberger, MJ.; Pope, ML. Minimal pairs test. Indiana University School of Medicine; Indianapolis, IN: 1988.
- Rubinstein JT, Miller CA. How do cochlear prostheses work? *Current Opinion in Neurobiology*. 1999; 9:399–404. [PubMed: 10448165]
- Seghal ST, Kirk KI, Hay-McCutcheon M. A comparison of children's familiarity with tokens on the PBK, LNT, and MLNT. *Annals of Otology, Rhinology, and Laryngology*. in press.
- Skinner MW, Clark GM, Whitford LA, Seligman PM, Staller SJ, Shipp DB, Shallop JK, Everingham C, Merapace CM, Ardnt PL, Antongenell T, Brimacombe JA, Pihl S, Daniels P, George CR,

- McDermott HJ, Beiter AL. Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. *American Journal of Otology*. 1994; 15:25–27.
- Skinner MW, Holden LK, Holden TA, Dowell RC, Seligman PM, Brimacombe JA, Beiter AL. Performance of postlingually deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant. *Ear and Hearing*. 1991; 12:3–22. [PubMed: 2026285]
- Sommers MS, Kirk KI, Pisoni DB. Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal hearing, and cochlear implant listeners I: The effects of response format. *Ear and Hearing*. 1997; 18:89–99. [PubMed: 9099558]
- Staller SJ, Beiter AL, Brimacombe JA. Use of the Nucleus 22 Channel Cochlear Implant System with children. *Volta Review*. 1994; 96:15–39.
- Svirsky MA. Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues. *Journal of the Acoustical Society of America*. 2000; 107:1521–1529. [PubMed: 10738806]
- Taft M, Forster KI. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14:638–647.
- Waltzman, SB.; Cohen, NL. Cochlear implants. Theime; New York: 2000.
- Waltzman SB, Cohen NL, Gomolin RH, Shapiro WH, Ozdamar SR, Hoffman RA. Long term results of early cochlear implantation in congenitally and prelingually deafened children. *American Journal of Otology*. 1994; 15(Suppl.):9–14. [PubMed: 8572107]
1. Haskins, HA. Unpublished Master's thesis. Northwestern University; 1949. A phonetically balanced test of speech discrimination for children..
 2. Broe, MB. Unpublished Ph.D. dissertation. University of Edinburgh; 1993. Specification theory: The treatment of redundancy in generative phonology..
 3. Frisch, SA. Unpublished Ph.D. dissertation. Northwestern University; 1996. Similarity and frequency in phonology..

Figure 1. Comparison of word (top panel) and phoneme (bottom panel) performance on the LNT easy word list. The MPEAK users are shown on the left; the SPEAK users are shown on the right. Error bars indicate a 95% confidence interval for the mean.

Figure 2. Comparison of word (top panel) and phoneme (bottom panel) performance on the LNT hard word list. The MPEAK users are shown on the left; the SPEAK users are shown on the right. Error bars indicate a 95% confidence interval for the mean.

Figure 3. Comparison of word (top panel) and phoneme (bottom panel) performance on the PBK. The MPEAK users are shown on the left; the SPEAK users are shown on the right. Error bars indicate a 95% confidence interval for the mean.

TABLE 1

Glossary of commonly used acroynms.

LNT	Lexical Neighborhoods Test: Test of open-set word recognition with “Easy” and “Hard” word lists that have different lexical characteristics.
MPEAK	Multipeak: Older speech processing strategy for Nucleus 22-channel cochlear implant that extracts linguistically relevant information from the signal.
NAM	Neighborhood Activation Model: Theoretical model of spoken word recognition in which words are activated by phonetic information and compete to be recognized.
PBK	Phonetically Balanced Kindergarten test: Test of open-set word recognition with four word lists that each use the same distribution of phonemes.
PCM	Phoneme Confusion Model: Theoretical model of spoken word recognition in which words are recognized only on the basis of recognizing individual phonemes.
SPAMR	Syllable Position Alignment for Matching and Retrieval: Theoretical model of spoken word recognition in which words are recognized by taking the results of phoneme recognition and finding the best matching lexical item.
SPEAK	Spectral Peak: Newer speech processing strategy for Nucleus 22-channel cochlear implant that acts as a filter bank.

TABLE 2

Phonemic contrasts in the Minimal Pairs Test.

Consonant Contrasts			Vowel Contrasts	
Place	Manner	Voicing	Vowel Place	Vowel Height
p/k	ʃ/t	p/b	ɪ/ə	i/ɔ
p/t	m/b	k/g	u/i	u/i ⁰
ʃ/f	f/p	v/f	u/ɪ	æ/i

TABLE 3

Number of participants simulated and mean and range for demographic variables and performance on behavioral tasks for MPEAK and SPEAK users. *t*-test indicates significant differences between groups.

	MPEAK	SPEAK	<i>t</i> -test
Descriptive statistics			
<i>N</i>	28	19	
Oral/total communication	12/17	11/8	
Chronological age	7.38 (3.95–10.4)	7.32 (4.45–10.8)	
Age at onset of SNHL	0.43 (0–1.8)	0.02 (0–0.4)	**
Age at implantation	5.6 (2.2–8.7)	5.5 (2.6–8.9)	
Mean pure tone average	113 dB (102–120)	108 dB (97–118)	*
Minimal Pairs Test			
Place	62% (35–75%)	70% (38–88%)	*
Manner	63% (41–88%)	81% (44–100%)	**
Voicing	57% (25–91%)	71% (44–100%)	**
Vowel place	81% (47–100%)	93% (69–100%)	**
Vowel height	79% (31–100%)	93% (75–100%)	**
Open-set word recognition			
LNT-Easy word list			
<i>N</i>	18	16	
Words correct	19% (0–50%)	52% (0–86%)	**
Phonemes correct	36% (0–65%)	67% (7–74%)	**
LNT-Hard word list			
<i>N</i>	13	15	
Words correct	13% (0–32%)	39% (16–76%)	**
Phonemes correct	35% (0–63%)	63% (39–87%)	**
PBK			
<i>N</i>	27	19	
Words correct	9% (0–30%)	28% (0–78%)	**
Phonemes correct	31% (1–68%)	57% (4–91%)	**

*
 $p < 0.05$.

**
 $p < 0.01$.

TABLE 4

Correlations between performance on Minimal Pairs Test and tests of open-set word recognition.

	MPEAK	SPEAK
LNT-Easy word list		
Words correct	+ 0.80 ^{**}	+0.78 ^{**}
Phonemes correct	+ 0.78 ^{**}	+ 0.71 ^{**}
LNT-Hard word list		
Words correct	+ 0.55 [*]	+0.72 ^{**}
Phonemes correct	+ 0.62 [*]	+0.67 ^{**}
PBK		
Words correct	+ 0.60 ^{**}	+0.69 ^{**}
Phonemes correct	+ 0.72 ^{**}	+0.72 ^{**}

*
 $p < 0.05$.^{**}
 $p < 0.01$.