RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 22 (1998)
Indiana University

Behavioral Data to Model Open-Set Word Recognition and Lexical Organization[1]

Stefan Frisch,[2] Ted A. Meyer,[3] David B. Pisoni,[3]
Mario A. Svirsky[3] and Karen I. Kirk[3]

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

---

[1] This work was supported by NIH-NIDCD Research Grant DC00064 and Training Grant DC00012 to Indiana University.

[2] Now at the Program in Linguistics, University of Michigan, Ann Arbor, MI.

[3] Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

S. FRISCH, T.A. MEYER, D.B. PISONI, M.A. SVIRSKY AND K.I. KIRK

Behavioral Data to Model Open-Set Word Recognition and Lexical Organization

Abstract. Children who acquire language through a cochlear implant (CI) provide a unique opportunity to study the physical and cognitive bases of speech perception, as the target language these children must acquire is the same as children without hearing loss, but the input to the nervous system from a cochlear implant is very different from normal hearing. Psycholinguistic models of speech perception in adults with normal hearing propose that the phonological lexicon is organized in a phonetic similarity space based on linguistic features and phonemes, and that these units are important for spoken word recognition. In this study, we examine software simulations of spoken word recognition using behavioral data on feature, phoneme, and word identification from children who use CIs. While the physical signal received through a CI is different from aurally perceived speech, we propose that the same abstract units of language are nonetheless a part of the linguistic knowledge of children who learn spoken language through CIs. Our results are consistent with the claim that children who use CIs recognize words in the context of perceptually similar words stored in a mental lexicon. In addition, while some feature distinctions are less adequately transmitted through their devices (e.g., place of articulation), children with CIs appear to utilize the available phonological contrasts to recognize spoken words.

Introduction

In this paper, we model performance by children with cochlear implants (CIs) on tests of open-set word recognition using their performance on closed-set tests of feature identification. Our model of word recognition does not employ any free parameters, and thus is highly constrained. Modeling the cognitive process of spoken word recognition based on feature identification performance is important to both theoretical and clinical research. The model of word recognition we propose is appropriate for both listeners who use CIs as well as those with normal hearing. It is thus a theoretical model of the process of spoken word recognition in general as well one way to address the question of whether the cognitive process of spoken word recognition is the same for both CI users and those with normal hearing. Comparing observed and predicted open-set word recognition performance from closed-set identification tasks may be particularly useful clinically for the pediatric CI population. If differences between observed and predicted performance can be attributed to particular components of the model, then it may be possible to provide intervention that is maximally beneficial to the child.

MODEL OF OPEN-SET WORD RECOGNITION

The model we employ simulates open-set word recognition with a three-step procedure, discussed in greater detail below. In the first step, a participant's closed-set feature identification performance ability is used to generate participant-specific phoneme confusion matrix. In the second step, the confusion matrix is applied to the string of phonemes in the target word (e.g., a stimulus item on a test of open-set word recognition) to produce a candidate target phoneme string based on that participant's feature identification abilities. In the third step, the target phoneme string is compared to the words in an on-line dictionary and the best matching lexical item is selected.

Our simulations used data obtained from children with cochlear implants observed at the Indiana University School of Medicine. In this population, feature identification performance is routinely assessed using the Minimal Pairs Test (Robbins, Renshaw, Miyamoto, Osberger & Pope, 1988). In this test, the participant is shown two pictures which are associated with words that differ by one phoneme (e.g., bear and pear). The difference in phonemes for each item in the test is a difference in one dimension of phonological contrast (e.g., place of articulation). Three different dimensions of contrast are tested for consonants (place, manner, and voicing) and two different dimensions of contrast are tested for vowels (vowel place and vowel height). If a particular dimension of contrast is not identifiable for some participant, that makes specific predictions about that listener's performance in word recognition. For example, if place of articulation cannot be reliably identified, then /p/, /t/, and /k/ are predicted to be confusable, as are /b/, /d/ and /g/, /m/, /n/, and /ŋ/, and so forth.

Performance on feature identification in the Minimal Pairs Test is used to generate predicted phoneme confusion matrices. First, closed-set feature identification performance is converted to predicted open-set feature identification performance by adjusting for the difference in chance performance associated with a closed-set task. For a two-alternative test like the Minimal Pairs Test, the equation is estimated open-set = 2·(observed closed-set - 50%). Second, we generate confusion matrices for each dimension of contrast by determining which phonemes are no longer distinct when a dimension of contrast is removed. Third, we combine these confusion matrices together depending on how reliably the dimensions of contrast are identified for a particular participant. The confusion matrix for each dimension of contrast in the combination is weighted by the participant's performance on the Minimal Pairs Test for that dimension. The weighted confusion matrices are added to create an overall predicted confusion matrix for each participant.

To simulate open-set spoken word recognition, the phoneme confusion matrix is applied to each phoneme in each word of a test of open-set word recognition. A model of word recognition which goes no further, i.e., which assumes that a word is recognized as the sum of its phonemes, will be referred to as the Phoneme Confusion Model (PCM [Frisch & Pisoni, 1998]). A model of spoken word recognition that is psychologically more plausible than PCM includes an additional step, where the mental lexicon is searched for a match to the target word. We use a simple procedure for lexical access that finds the word in the lexicon with the greatest number of matching phonemes to the target word. Overlap is determined based on the optimal alignment of syllabic positions between two words. For example, the words hello and heavy are best aligned at their beginnings, and they overlap in the two phonemes /h/ and /ɛ/. The words hello and low are best aligned by matching the second syllable of hello with low. Again, they overlap by two phonemes, in this case /l/ and /o/. If two words tie for the greatest number of matching phonemes, the tie is broken by choosing the word with the highest frequency of use in the language. The model of word recognition that includes lexical access is referred to as Syllable Position Alignment for Matching and Retrieval (SPAMR).

The SPAMR model uses a simulated child's mental lexicon containing a subset of words from an on-line version of Webster's pocket dictionary that is used as an approximation of an adult's mental lexicon (Nusbaum, Pisoni & Davis, 1984). The simulated child's lexicon was generated by using high usage frequency words from the adult lexicon, plus all of the words on the two tests of open-set word recognition used in these simulations, the Lexical Neighborhoods Test (LNT [Kirk, Pisoni & Osberger, 1995]) and the Phonetically Balanced Kindergarten test (PBK[Haskins]). Thus, we assume all of the words on the word recognition tests are known by the children. It had been claimed that many of the words on the PBK are unfamiliar to young children, so this assumption may be incorrect (Kirk, Pisoni & Osberger, 1995; Sehgal, Kirk & Hay-McCutcheon, 1998). The simulated child's lexicon contained about 1400 words.

DATA

Behavioral data from 30 pediatric CI users with Nucleus 22 channel CIs and either the SPEAK (Skinner et al., 1991) or MPEAK (Skinner et al., 1994) processing strategies were used. Test scores were from 1.5-2.0 years post implant. Eighteen children used MPEAK and 12 children used SPEAK. The children had mean age of onset of profound hearing loss of 0.32 years (median 0), mean age of implantation of 4.45 years (median 4.60). The children were evenly divided between oral and total communication programs.

Our simulations used observed data and simulated recognition of the words in the Minimal Pairs Test, LNT, and the PBK. The LNT contains two sets of stimuli, called 'easy' and 'hard', based on their confusability with other words in the lexicon. 'Easy' words are high usage frequency words that are similar to a just few other words which are of low usage frequency. 'Hard' words are low usage frequency words that are similar to many high usage frequency words. Stimuli on the PBK vary greatly in their lexical characteristics, as the PBK lists were balanced for their phonemic, not lexical, properties.

RESULTS

Comparison of observed and predicted performance averaged across all the children is given in Figure 1. Observed and predicted performance is shown for the LNT easy and hard words separately. First, note that the model with a lexical access component (SPAMR) makes a very good prediction of observed performance in words correct on both easy and hard words on the LNT. The SPAMR model predicts performance in words correct much better than the PCM. The difference between PCM and SPAMR is greatest for the LNT easy words. These are words which are in sparse neighborhoods of the mental lexicon, so lexical knowledge can be used to 'fill in' incomplete acoustic phonetic information. For the PBK, the model with lexical knowledge performed significantly above the observed level, while the model with no lexicon was more accurate. This pattern supports the claim that young children may be unfamiliar with many of the words on the PBK (Kirk, Pisoni & Osberger, 1995). Since predictions for words correct on the LNT improved with the addition of a stage of lexical access, we believe modeling lexical access is a necessary part of modeling open-set spoken word recognition with these pediatric CI users. Overall, these simulations suggest that the process of spoken word recognition is similar for both pediatric CI users and adults with normal hearing.

Insert Figure 1 about here

Comparing observed and predicted performance in phonemes correct shows that both PCM and SPAMR make predictions that are very close to observed performance, only over-predicting for the PBK. Based on the good predictions for phonemes correct, we conclude that closed-set feature identification can successfully predict phoneme identification in an open-set word recognition task. But note that phoneme prediction alone, as in the PCM, is not sufficient to account for word recognition. Lexical access is an equally important component of the process.

CONCLUSION

We have demonstrated significant progress in predicting open-set word recognition performance from closed-set feature identification performance in children who use cochlear implants. Different aspects of the model account for phoneme identification, and lexical access and word recognition, so comparing observed and predicted open-set word recognition performance may be useful clinically for

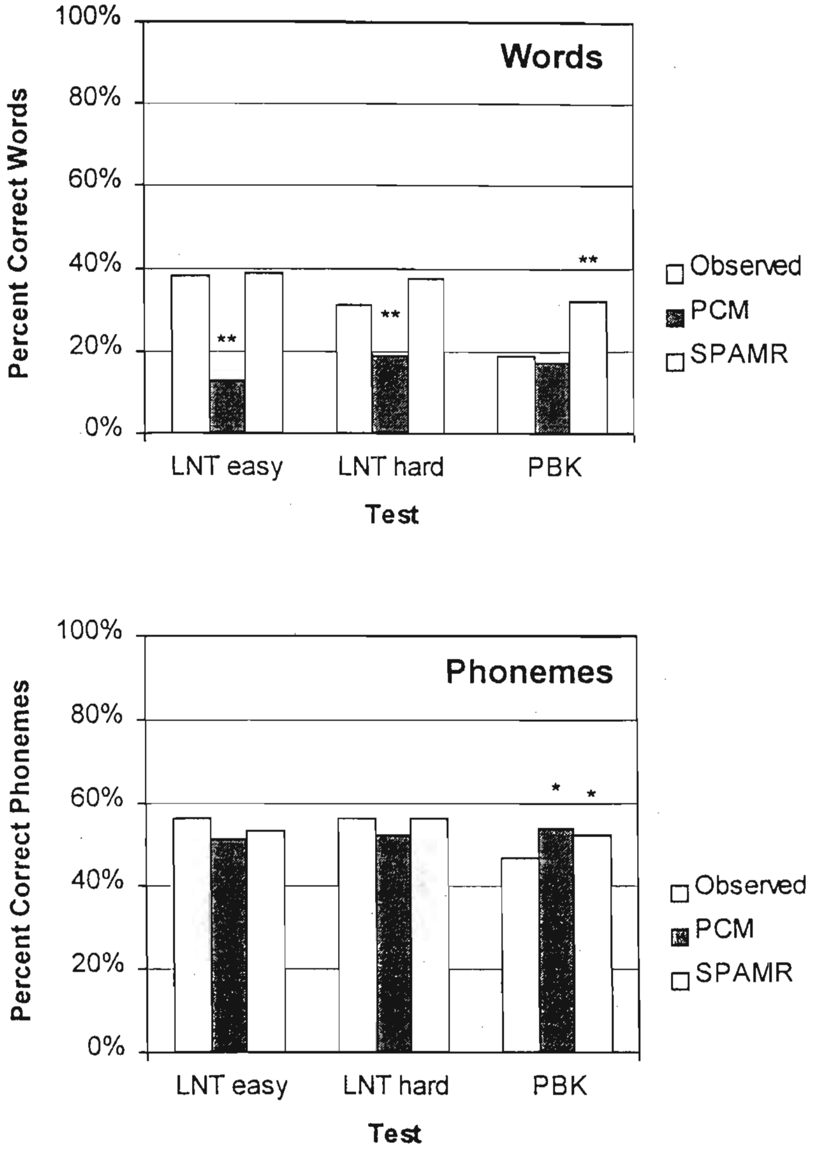

Figure 1. Observed and predicted performance for LNT easy words, LNT hard words, and the PBK words. The top panel shows performance scored in percent words correct. The bottom panel shows performance scored in percent phonemes correct. Significant differences between observed and predicted performance on a paired t-test are indicated for the models: $^*p < 0.05$, $^{**}p < 0.01$.

highlighting aspects of performance which would be most crucial to foster further development and improvement in spoken language ability. We have shown how modeling can be used to predict young children's potential performance on open-set word recognition tests. However, the psycholinguistic model contains no mechanisms (apart from the choice of lexicon) specific to children who use CIs. Thus, this model may be useful for other clinical populations or children and adults with normal hearing. The generality of the model provides evidence that the process of spoken word recognition by children with CIs may reflect similar units of linguistic structure and lexical organization evident in the process of spoken word recognition by adults with normal hearing.

References

Frisch S, & Pisoni DB. (1998). Predicting spoken word recognition performance from feature discrimination scores in pediatric cochlear implant users: A preliminary analysis. *Research on Spoken Language Processing Progress Report 21*. Bloomington, IN: Speech Research Laboratory, Indiana University. Pp. 261-288.

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

Kirk KI, Pisoni DB, & Osberger M. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing, 16,* 470-481.

Nusbaum HC, Pisoni DB, & Davis CK. (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. *Research on Spoken Language Processing Progress Report 10*. Bloomington, IN: Indiana University. Pp. 357-376.

Robbins AM, Renshaw JJ, Miyamoto RT, Osberger MJ, & Pope ML. (1988). *Minimal pairs test.* Indianapolis, IN: Indiana University School of Medicine.

Sehgal ST, Kirk KI, & Hay-McCutcheon M. (1998).A comparison of children's familiarity with tokens on the PB-K, LNT and MLNT. This volume.

Skinner MW, Clark GM, Whitford LA, Seligman PM, Staller SJ, Shipp DB, Shallop JK, Everingham C, Menapace CM, Arndt PL, Antogenelli T, Brimacombe JA, Pijl S, Daniels P, George CR, McDermott HJ, & Beiter AL. (994). Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. *American Journal of Otology, 15,* 25-27.

Skinner MW, Holden LK, Holden TA, Dowell RC, Seligman PM, Brimacombe JA, & Beiter AL. (1991). Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucelus multi-electrode cochlear implant. *Ear and Hearing, 12,* 3-22.