Published in final edited form as:

Ann Otol Rhinol Laryngol Suppl. 2000 December ; 185: 60–62.

USING BEHAVIORAL DATA TO MODEL OPEN-SET WORD RECOGNITION AND LEXICAL ORGANIZATION BY PEDIATRIC COCHLEAR IMPLANT USERS

Stefan Frisch, PhD, Ted A. Meyer, MD, PhD, David B. Pisoni, PhD, Mario A. Svirsky, PhD, and Karen Iler Kirk, PhD

Department of Psychology, Indiana University, Bloomington (Frisch, Pisoni), and the Department of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, Indianapolis (Meyer, Pisoni, Svirsky, Iler Kirk), Indiana

INTRODUCTION

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 for children without hearing loss, but the input to the nervous system from a CI 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. Although 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.

In this paper, we model performance by children with 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 therefore is highly constrained. Modeling the cognitive process of spoken word recognition on the basis of 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 and those with normal hearing. It is therefore 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

Our simulations used data from the pediatric CI user population seen at the Indiana University School of Medicine. In this population, feature identification performance is routinely assessed with the Minimal Pairs Test.¹ In this test, the participant is shown 2 pictures that are associated with words that differ on 1 phoneme (eg, *bear* and *pear*). The difference in phonemes for each item in the test is a difference in 1 dimension of phonological contrast (eg, manner of articulation). Three different dimensions of contrast are tested for consonants (place, manner, and voicing), and 2 different dimensions of contrast are tested for vowels (vowel place and vowel height). If a particular dimension of contrast is

not identifiable for a participant, specific predictions about that listener's performance in word recognition can be made. For example, if manner of articulation cannot be reliably identified, then /b/, /v/, and /m/ are predicted to be confusable, as are /p/ and /f/, /t/ and /s/, 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 base rate associated with a closed-set task. For a 2-alternative test like the Minimal Pairs Test, the equation is Estimated open-set = $2 \times (\text{observed closed-set} - 50\%)$. 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 depending on how reliably the dimensions of contrast are identified by a particular listener. The confusion matrix for each dimension of contrast in the combination is weighted by the listener'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 listener.

The predicted confusion matrix is applied to the string of phonemes in the target word (eg, a stimulus item on a test of open-set word recognition) to produce a response phoneme string based on that participant's feature identification abilities. The response phoneme string is compared with the words in an on-line dictionary, and the best matching lexical item is selected. For example, if the target word is "cat" /kat/, the model might respond "cat," or it could respond "bat," "dat," "cap," "guk," or any of a number of other responses. Of the responses listed, "cat," "bat," and "cap" are words in the English language, whereas "dat" and "guk" are nonsense syllables.

To simulate open-set spoken word recognition, the phoneme confusion matrix is applied to each phoneme in each word of an open-set word recognition test. A model of word recognition that goes no further, ie, that assumes that a word is recognized as the sum of its phonemes, will be referred to as the Phoneme Confusion Model, or PCM.² A model of spoken word recognition that is psychologically more plausible than PCM includes an additional step in which 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 phonemes matching the target word. Overlap is determined on the basis of the optimal alignment of syllabic positions between 2 words. For example, the words *hello* and *heavy* are best aligned at their beginnings, and they overlap in the 2 phonemes /h/ and /E/. The words *hello* and *low* are best aligned by matching the second syllable of *hello* with *low*. Again, they overlap by 2 phonemes, in this case /l/ and /o/. If 2 words tie for the greatest number of matching phonemes, the tie is broken by choosing the word with the highest usage frequency. The model of word recognition that includes lexical access is referred to as Syllable Position Alignment for Matching and Retrieval, or SPAMR.

The SPAMR model uses a simulation of a child's mental lexicon containing a subset of words from an on-line version of a Webster's pocket dictionary that is an approximation of an adult's mental lexicon.³ The simulated child's lexicon was generated with high-frequency words from the adult lexicon, plus all of the words on the 2 tests of open-set word recognition used in these simulations, the Lexical Neighborhood Test (LNT)⁴ and the Phonetically Balanced–Kindergarten test (PBK).⁵ Therefore, we assume all of the words on the word recognition tests are known by the children. It has been claimed that many of the words on the PBK are unfamiliar to young children, so the previous assumption may be incorrect.⁴ The simulated child's lexicon contained about 1,400 words.

DATA

Behavioral data from 30 pediatric CI users with Nucleus 22-channel CIs and either the spectral peak (SPEAK)⁶ or multipeak (MPEAK)⁷ processing strategies were used. The test scores were obtained 1.5 to 2.0 years after implantation. There were 18 children who used MPEAK and 12 children who used SPEAK. The children had a mean age at onset of profound hearing loss of 0.32 years (median, 0) and a mean age at implantation of 4.45 years (median, 4.60). The children were evenly divided between those who use mainly oral communication and those who use total communication, a combination of spoken and manual communication. Because the data set was relatively small and in the present study we examined only averaged data, the data for the children using the different processing strategies as well as different modes of communication (oral and total communication) were analyzed together.

Our simulations used observed data and simulated recognition of the words in the Minimal Pairs Test, the LNT, and the PBK. The LNT contains 2 sets of stimuli, called “easy” and “hard” on the basis of their confusability with other words in the lexicon. “Easy” words are high-usage frequency words that are similar to only a 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. The stimuli on the PBK vary greatly in their lexical characteristics, because the PBK lists were balanced for their phonemic, not lexical, properties.

RESULTS

A comparison of observed and predicted performance averaged across all the children is given in the Figure. The observed and predicted performances are 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 does the PCM. The difference between PCM and SPAMR is greatest for the LNT easy words. These are words that 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 better than the observed level, whereas the model with no lexicon was more accurate. This pattern supports the claim that young children are unfamiliar with many of the words on the PBK.⁴ Because 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.

Comparing observed and predicted performance in phonemes correct shows that both PCM and SPAMR make predictions that are very close to observed performance, only overpredicting for the PBK. On the basis of 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.

CONCLUSIONS

We have demonstrated significant progress in predicting open-set word recognition performance from closed-set feature identification performance in children who use CIs. Different aspects of the model account for phoneme identification and lexical access in word recognition. Comparing observed and predicted open-set word recognition performance may be useful clinically, because closed-set tests have a more straightforward response format

and therefore may be more developmentally appropriate for very young children. However, the psycholinguistic model contains no mechanisms (apart from the choice of a lexicon) specific to children who use CIs. Therefore this model may be useful for other clinical populations, including children and adults with normal hearing. The model will also be used to predict performance for individual CI users who employ different modes of communication. The generality of the model provides evidence that the process of spoken word recognition by children with CIs reflects units of linguistic structure and lexical organization similar to those evident in the process of spoken word recognition by adults with normal hearing.

Acknowledgments

This work was supported by National Institutes of Health/National Institute on Deafness and Other Communication Disorders grants DC00064 and DC00012.

References

1. Robbins, AM.; Renshaw, JJ.; Miyamoto, RT.; Osberger, MJ.; Pope, ML. Minimal Pairs Test. Indianapolis, Ind: Indiana University School of Medicine; 1988.
2. Frisch, S.; Pisoni, DB. Research on spoken language processing: progress report 21. Bloomington, Ind: Speech Research Laboratory, Indiana University; 1997. Predicting spoken word recognition performance from feature discrimination scores in pediatric cochlear implant users: a preliminary analysis; p. 261-88.
3. Nusbaum, HC.; Pisoni, DB.; Davis, CK. Research on spoken language processing: progress report 10. Bloomington, Ind: Indiana University; 1984. Sizing up the Hoosier Mental Lexicon: measuring the familiarity of 20,000 words; p. 357-76.
4. Kirk KI, Pisoni DB, Osberger M. Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear Hear.* 1995; 16:470–81. [PubMed: 8654902]
5. Haskins, H. A phonetically balanced test of speech discrimination for children [Thesis]. Evanston, III: Northwestern University; 1949.
6. Skinner MW, Holden LK, Holden TA, et al. Performance of post-linguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant. *Ear Hear.* 1991; 12:3–22. [PubMed: 2026285]
7. Skinner MW, Clark GM, Whitford LA, et al. Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. *Am J Otol.* 1994; 15:25–7.

Figure 1.

Figure Observed and predicted performance for Lexical Neighborhood Test (LNT) easy words, LNT hard words, and Phonetically Balanced-Kindergarten (PBK) words. PCM — Phoneme Confusion Model. SPAMR — Syllable Position Alignment for Matching and Retrieval. A) Performance scored in percent words correct. B) Performance scored in percent phonemes correct. Significant differences between observed and predicted performances on paired *t*-tests are indicated for models: * — $p < .05$; ** — $p < .01$.