Published in final edited form as:

Ann Otol Rhinol Laryngol Suppl. 2000 December ; 185: 68–70.

MODELING PHONEME AND OPEN-SET WORD RECOGNITION BY COCHLEAR IMPLANT USERS: A PRELIMINARY REPORT

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

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

INTRODUCTION

The overall goal of modeling speech perception by cochlear implant (CI) users is to understand the sensory and perceptual mechanisms that CI users employ to understand spoken language. Frisch and Pisoni¹ (also Frisch et al, this supplement, pp 60–62) have developed mathematical models to predict open-set phoneme and word recognition from closed-set feature identification scores for a group of children with profound prelingual hearing losses who use CIs. In the models, individual phonemes are characterized by their various phonetic features. Probabilities of identification of various phonetic features are obtained from vowel and consonant confusion matrices and are used as input parameters to 2 variations of the model. In the first model, candidate responses in an open-set word recognition test are generated by stringing together the most likely phonemes on the basis of feature identification scores. This version, referred to as the Phoneme Confusion Model (PCM),¹ assumes that a word is recognized by identifying each phoneme independently of the others. A model of open-set word recognition that is psychologically more plausible than PCM includes an additional step, in which a lexicon is searched for a possible match to the model output. This model is referred to as Syllable Position Alignment for Matching and Retrieval (SPAMR; this supplement, pp 60–62). The SPAMR model uses an on-line version of a Webster's pocket dictionary² as an approximation of the mental lexicon, and it finds the word in the lexicon with the greatest number of phonemes matching the model output. In the present study, we obtained predictions of open-set spoken word recognition from averaged vowel and consonant confusion matrices from a group of adult CI users.

METHODS

Data

Vowel and consonant confusion matrices³ obtained from 7 adults who were experienced users of the Nucleus 22 CI with the spectral peak (SPEAK) strategy⁴ were used as input to the PCM and SPAMR models. Although the confusion matrices for the individual CI users were available at the time of the study, each CI user heard only 12 repetitions of each stimulus. The average word and phoneme scores on the consonant-nucleus-consonant (CNC) test⁵ for the 7 CI users were compared with the predicted scores from the outputs of the PCM and SPAMR models.

Model

The average vowel and consonant confusion matrices were input to the PCM or SPAMR models, which generate strings of phonemes as possible responses to the stimulus. Given a particular word as input, the models output different strings of phonemes with the proper statistical distribution for the identification of different phoneme features. For example, if

the stimulus were the word “seek” /sik/, the model would output/sik/ a certain number of times depending on the feature identification scores. The models might also output “zut,” “fak,” “seat” /sit/, or “sick” /sIk/ as possible responses that are similar to the target word. The PCM model would simply keep track of the number of correct phonemes and words. The SPAMR model, however, would find “seat” /sit/or “sick” /sIk/ in the lexicon, and the number of phonemes would be counted, but it would not find “zut” or” fak.” The SPAMR model would search the lexicon for the word or words closest in phonemic structure to “zut” or “fak” and consider these the responses. The numbers of words and phonemes correct were calculated from the model output and compared with the numbers of words and phonemes obtained by the CI users.

RESULTS

A comparison of the average observed and predicted performances on the CNC test is given in the Figure. First, note that the PCM model, which does not use a lexicon, underpredicts the observed performance in words correct. The PCM model predicted only 40% words correct, whereas the listeners actually obtained approximately 51 % words correct. The SPAMR model, which includes a lexicon, obtained 55% word identification and 80% phoneme identification, which was quite close to the observed CNC scores for words (51 %) and phonemes (76%) correct, and thus provides a better prediction of observed performance than the PCM model. Comparing observed and predicted performance in *phonemes* correct shows that both models make predictions that are very close to observed performance.

SUMMARY

On the basis of the good predictions for phonemes correct, we conclude that closed-set feature identification may successfully predict phoneme identification in an open-set word recognition task. For word recognition, however, the PCM model underpredicted observed performance, and the addition of a mental lexicon (ie, the SPAMR model) was needed for a good match to data averaged across 7 adults with CIs. The predictions for words correct improved with the addition of a lexicon, providing support for the hypothesis that lexical information is used in open-set spoken word recognition by CI users. The perception of words more complex than CNCs is also likely to require lexical knowledge (Frisch et al, this supplement, pp 60–62). In the future, we will use the performance of individual CI users on psychophysical tasks to generate predicted vowel and consonant confusion matrices to be used to predict open-set spoken word recognition.^{6,7}

Acknowledgments

This study was supported by grants from the National Institutes of Health/National Institute on Deafness and Other Communication Disorders (DC00064, T32 DC00012), the Deafness Research Foundation, and the National Organization for Hearing Research.

We thank Dr Margaret W. Skinner for allowing us access to the vowel and consonant confusion matrices.

References

1. Frisch, S.; Pisoni, DB. Research on spoken language processing: progress report. Vol. 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.
2. Nusbaum, HC.; Pisoni, DB.; Davis, CK. Research on spoken language processing: progress report. Vol. 10. Bloomington, Ind: Indiana University; 1984. Sizing up the Hoosier Mental Lexicon: measuring the familiarity of 20,000 words; p. 357-76.

3. Skinner MW, Holden LK, Holden TA. Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy. *Ann Otol Rhinol Laryngol Suppl.* 1995; 104(suppl 166):307–11. [PubMed: 7668683]
4. 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; 15:15–27. [PubMed: 8572106]
5. Peterson G, Lehiste I. Revised CNC lists for auditory tests. *J Speech Hear Disord.* 1962; 27:62–70. [PubMed: 14485785]
6. Svirsky MA, Meyer TA. A mathematical model of vowel perception by cochlear implantees who use the SPEAK stimulation strategy. *Association for Research in Otolaryngology.* 1997; 20:abstract 234.
7. Meyer TA, Svirsky MA, Castor MB, Falsafi S, Simmons PM. A psychophysically-based model of consonant perception by multi-channel cochlear implant users [Abstract]. *J Acoust Soc Am.* 1997; 102:3187.

Figure. Observed and predicted performance for CNC words. Percent words correct is shown in left portion of Figure, and percent phonemes correct is shown in right. PCM — Phoneme Confusion Model. SPAMR — Syllable Position Alignment for Matching and Retrieval.