

RESEARCH ON SPOKEN LANGUAGE PROCESSING

Progress Report No. 22 (1998)

Indiana University

**Modeling Phoneme and Open-Set Word Recognition
by Cochlear Implant Users Based on Psychophysical Performance:
A Preliminary Report¹**

Ted A. Meyer,² Stefan Frisch,³ Mario A. Svirsky² and David B. Pisoni²

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

¹ This study was supported by grants from NIH/NIDCD DC00064, T32 DC00012, DRF, and NOHR to Indiana University. We would like to thank Dr. Margaret W. Skinner for allowing us access to the vowel and consonant confusion matrices.

² DeVault Otologic Research Laboratory, Department of Otolaryngology-Head & Neck Surgery, Indiana University school of Medicine, Indianapolis, IN.

³ Program in Linguistics, University of Michigan, Ann Arbor, MI.

Modeling Phoneme and Open-Set Word Recognition by Cochlear Implant Users Based on Psychophysical Performance: A Preliminary Report

Abstract. We have recently made some progress in our ability to explain perceptual performance by adult cochlear implant (CI) users, and this research may eventually be applicable to children. Quantitative, psychophysically-based models of vowel (Svirsky & Meyer, 1997) and consonant (Meyer, Svirsky, Castor, Falsafi & Simmons, 1997) perception generated predicted confusion matrices that were similar to matrices averaged across seven adult CI users (Skinner, Holden & Holden, 1995). Frisch and colleagues (Frisch & Pisoni, 1997; Frisch, Meyer, Pisoni, Svirsky & Kirk, submitted) have developed models to predict open-set word recognition by CI users based on their vowel and consonant feature recognition. Using the Frisch et al. model, we obtained predictions of open-set spoken word recognition from observed vowel and consonant confusion matrices for a group of adult CI users, and we compared these predictions to observed performance on a test of open-set spoken word recognition. In the future, we will use both models to predict open-set spoken word recognition from performance on psychophysical tests for individual CI users, including children.

Introduction

The overall goal of this project is to enhance our knowledge of the underlying sensory and perceptual mechanisms used by cochlear implant (CI) users to understand spoken language. The traditional approach has been to perform correlational analyses between the different psychophysical and speech perception performance measures by CI users. Correlations from 0.6 to 0.8 have been found between speech perception and psychophysical performance that assess a CI user's ability to make fine discriminations in the temporal domain (Collins, Wakefield & Feinman, 1994). Such correlations, however, cannot explain the mechanisms CI users employ to identify speech sounds.

In contrast, our proposed modeling approach provides a detailed mathematical description of how CI users employ the information they receive through their devices to perceive speech. Moreover, the predictions provided by the model are not simply predicted speech perception scores, but entire confusion matrices. The model makes specific predictions as to which pairs of sounds should be more easily confused, or not confused, by a CI user, and it allows us to examine many of the complexities inherent in phoneme recognition. The underlying dimensions in this model are similar to dimensions suggested to be important for vowel and consonant recognition with a CI (Dorman, Dankowski, McCandless, Parkin & Smith, 1991). The model of vowel perception predicts that a CI user's ability to discriminate vowels is directly related to the ability of the CI user to discriminate differences in the place of stimulation along the cochlea. The model of consonant perception predicts that a CI user's ability to discriminate consonants is related to the ability of the CI user to discriminate differences in the place of stimulation along the cochlea as well as the duration of a silent period in the stimulus and the ratio of the amplitude of the stimulus above and below 800 Hz.

To date, this approach has been successful in modeling vowel and consonant confusions for a group of adult CI users. The model of vowel perception predicted performance of 86.5% (84.5% observed) with an RMS error between the two confusion matrices of 5%, while the model of consonant perception predicted performance of 78.2% (77.2% observed) with an RMS error of 5%.

Model of Open-Set Word Recognition

Recently, Frisch and colleagues 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 CI's (Frisch & Pisoni, 1997; Frisch et al., submitted). In the models, individual phonemes are characterized by their various features which allows for identification of the different phonemes. Probabilities of identification of various phonetic features are obtained from vowel and consonant confusion matrices and are used as input parameters to the model. The model simulates open-set word recognition with a two step procedure. In the first step, probabilities from phoneme confusion matrices are 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 response (another string of phonemes) based on that patient's feature identification abilities. This version, referred to as the Phoneme Confusion Model (PCM[Frisch & Pisoni, 1997]), assumes that a word is recognized as the sum of its phonemes. A model of spoken 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[Frisch et al., submitted]). The SPAMR model uses an on-line version of Webster's pocket dictionary (Nusbaum, Pisoni & Davis, 1984) as an approximation of the mental lexicon, and it finds the word in the lexicon with the greatest number of matching phonemes to the model output. As a first attempt at predicting open-set spoken word recognition from psychophysical performance, we predicted open-set word recognition from observed vowel and consonant confusion matrices.

Methods

Data

We used vowel and consonant confusion matrices as well as word and phoneme scores on the CNC test (Peterson & Lehiste, 1962) averaged across seven adults who were experienced users of the Nucleus-22 CI with the SPEAK (Skinner et al., 1994) strategy (Skinner et al., 1995).

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 percentage of correct feature identification scores. The model might also output "seat" /sit/ as well as "sick" /sik/ as possible responses that are similar to the target word. The number of words and phonemes correct were calculated from the model output.

Results

Comparison of average observed and predicted performance on the CNC Test is given in Figure 1. First, note that without a lexicon, the PCM model underpredicts observed performance in words correct. The PCM model predicted only 40% words correct while 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 which are very close to observed performance.

 Insert Figure 1 about here

Summary

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. For word recognition, however, the PCM model underpredicted observed performance, and the addition of a mental lexicon (i.e., the SPAMR model) was needed for a good match to data averaged across 7 adults with CI's. 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 CNC's is also likely to require lexical knowledge (Frisch et al., submitted). In the future, we will use performance by individual CI users on psychophysical tasks to generate predicted vowel and consonant confusion matrices to be used to predict open-set spoken word recognition.

References

- Collins LM, Wakefield GH, & Feinman GR. (1994). Temporal pattern discrimination and speech recognition under electrical stimulation. *Journal of the Acoustical Society of America*, 96, 2731-2737.
- Dorman MF, Dankowski K, McCandless G, Parkin JL, & Smith L. (1991). Vowel and consonant recognition with the aid of a multichannel cochlear implant. *Quarterly Journal of Experimental Psychology*, 43, 585-601.
- Frisch S, & Pisoni DB. (1997). Predicting spoken word recognition performance from feature discrimination scores in pediatric cochlear implant users: A preliminary analysis. In *Research on Spoken Language Processing Progress Report 21*(pp. 261-288). Bloomington, IN: Speech Research Laboratory, Indiana University.
- Frisch S, Meyer TA, Pisoni DB, Svirsky MA, & Kirk KI. (Submitted). Using behavioral data to model open-set word recognition and lexical organization by pediatric cochlear implant users. *Annals of Otology, Rhinology and Laryngology*.
- Meyer TA, Svirsky MA, Castor MB, Falsafi S, & Simmons PM. (1997). A psychophysically-based model of consonant perception by multi-channel cochlear implant users. *Journal of the Acoustical Society of America*, 10, 3187 (Abstract).
- Nusbaum HC, Pisoni DB, & Davis CK. (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. In *Research on Spoken Language Processing: Progress Report 10* (pp. 357-376). Bloomington, IN: Indiana University.
- Peterson G, & Lehiste I. (1962). Revised CNC lists for auditory tests. *Journal of Speech and Hearing Disorders*, 27, 62-70.

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. (1994). Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. *American Journal of Otology*, 15, 15-27.

Skinner MW, Holden LK, & Holden TA. (1995). Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy. *Annals of Otology, Rhinology and Laryngology*, 95;104 (Suppl. 166), 307-311.

Svirsky MA, & Meyer TA. (1997). A mathematical model of vowel perception by cochlear implantees who use the SPEAK stimulation strategy. *ARO Abstracts*, 20: No. 234 (Abstract).