

THE PHONEME AND THE GEOMETRY OF DECISION REGIONS IN SPEECH PERCEPTION

Terrance M. Nearey

Department of Linguistics, University of Alberta, Edmonton, AB T6G 2E7

INTRODUCTION

Various architectures have been proposed to account for human speech perception. This paper focuses on a non-parametric method for comparing them. Comparison of distinct architectures is not as easy as it may seem. One basic distinction hinges on whether a system is interactive, (i.e., has feedback between levels). Massaro (1989, *Cognitive Psych.* 21, 398-421) demonstrated that certain predictions of the TRACE model were not compatible with empirical results, while those from his feed-forward fuzzy logical models were. McClelland (1991 *Cognitive Psych.* 23, 1-44) showed that the failings of TRACE and of a simpler interactive activation model (IAC) may be due to the choice model employed and not to the presence of feedback.

In experiments where more than one stimulus property is varied, simple geometric properties of decision regions may differentiate perceptual models. This criterion has previously been used to help characterize differences in computational power among neural networks with zero, one or two hidden layers by Lippmann (1987, *IEEE ASSP Magazine* 4, 266-278). For evaluation of perceptual models, decision regions are defined by the modal response category, i.e., the category that is predicted to have the maximum number of responses on repeated presentation of the same stimulus. Such a characterization is independent of such complex issues as the choice process and the location of noise sources.

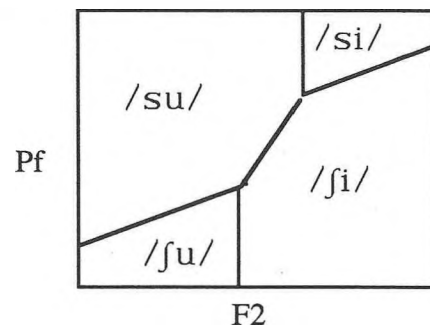
2 A CASE STUDY

Consider a problem in speech perception involving the words 'see, sue, she, shoe.' These may be synthesized by varying two stimulus parameters. 1) Pf, the frequency of a fricative pole. Lowering Pf can change an isolated /s/-sound to a /ʃ/-sound. 2) F2, the frequency of the second formant of the vocalic part of the syllable. Lowering F2 changes an isolated /i/ to /u/. It is well known (Whalen, 1989, *Percep. and Psychophys.*, 46, 284-292) that the boundary between /s/ and /ʃ/ on the Pf axis is lower when the following vowel is /u/ rather than /i/. Furthermore, the boundary between /i/ and /u/ is lower on the F2 axis when the preceding vowel is changed from /s/ to /ʃ/.

There are several ways such shifts might occur. One would be through an interactive-activation (IAC) model where the output of consonant units feeds back to adjacent vowel units and output from vowel units feeds back to adjacent consonant units. Results at least partly compatible with those described in the previous paragraph could be

obtained if /i/ output fed back with a larger positive weight to the input of /ʃ/, while /u/ output fed back more positively to /s/ input. Similarly, /s/ output could feed back more positively to /u/, and /ʃ/ output more positively to /i/.

However, the apparent compatibility breaks down in detail. Several purely feed-forward (non-interactive) alternatives were examined by Nearey (1990, *J. Phonetics* 18, 347-373) to model the results of the experiment of Whalen (1989) involving a two dimensional (Pf, F2) continuum. The results of this analysis indicated that the most appropriate model had decision regions depicted schematically:



The line separating /su/ from /ʃu/ is parallel to that separating /si/ from /ʃi/. These lines are not perpendicular to either axis. The boundaries of these regions are simultaneously modified in a continuous fashion by both stimulus properties. The /s/ response regions are favored by either increasing Pf or by decreasing F2. This is also grossly compatible with an IAC model. For example, higher F2 favors /i/ responses and the output of the /i/ unit feeds back more strongly to the input of the /ʃ/ unit, so raising F2 continuously (assuming output units have not saturated) increases /ʃ/ response area continuously.

The situation is quite different for the vowel boundaries. Both the /su/ - /ʃu/ boundary and the /si/-/ʃi/ boundary are unaffected by the frequency of Pf. The vowel choices are affected by the consonant choices by only what amounts to a bias effect. The /ʃi/ region is relatively larger than the /ʃu/ region, while the /su/ region is larger than the /ʃu/ region, but the shift in preference *does not vary continuously* with the frequency of Pf. However, the activation levels of the fricative units, are varying continuously and the usual richly interconnected architecture of IAC or TRACE would predict continuous change in vowel activations because of feedback.

Additional examples of applications of decision region geometry to issues in perceptual models will be discussed.