COGNITIVE PSYCHOLOGY 18, 1–86 (1986) 0010-0285/86 $7.50

The TRACE Model of Speech Perception

JAMES L. MCCLELLAND

Carnegie–Mellon University

AND

JEFFREY L. ELMAN

University of California, San Diego

We describe a model called the TRACE model of speech perception. The model is based on the principles of interactive activation. Information processing takes place through the excitatory and inhibitory interactions of a large number of simple processing units, each working continuously to update its own activation on the basis of the activations of other units to which it is connected. The model is called the TRACE model because the network of units forms a dynamic processing structure called "the Trace," which serves at once as the perceptual processing mechanism and as the system's working memory. The model is instantiated in two simulation programs. TRACE I, described in detail elsewhere, deals with short segments of real speech, and suggests a mechanism for coping with the fact that the cues to the identity of phonemes vary as a function of context. TRACE II, the focus of this article, simulates a large number of empirical findings on the perception of phonemes and words and on the interactions of phoneme and word perception. At the phoneme level, TRACE II simulates the influence of lexical information on the identification of phonemes and accounts for the fact that lexical effects are found under certain conditions but not others. The model also shows how knowledge of phonological constraints can be embodied in particular lexical items but can still be used to influence processing of novel, nonword utterances. The model also exhibits categorical perception and the ability to trade cues off against each other in phoneme identification. At the word level, the model captures the major positive feature of Marslen-Wilson's COHORT model of speech perception, in that it shows immediate sensitivity to information favoring one word or set of words over others. At the same time, it overcomes a difficulty with the COHORT model: it can recover from underspecification or mispronunciation of a word's beginning. TRACE II also uses lexical information to segment a stream of speech into a sequence of words and to find word beginnings and endings, and it simulates a number of recent findings related to these points. The TRACE model has some limitations, but we believe it is a step toward a psychologically and computationally adequate model of the process of speech perception. © 1986 Academic Press, Inc.

The work reported here was supported in part by a contract from the Office of Naval Research (N-00014-82-C-0374), in part by a grant from the National Science Foundation (BNS-79-24062), and in part by a Research Scientists Career Development Award to the first author from the National Institute of Mental Health (5-K01-MH00385). We thank Dr. Joanne Miller for a very useful discussion which inspired us to write this article in its present form. David Pisoni was extremely helpful in making us deal more fully with several important issues, and in alerting us to a large number of useful papers in the literature. We also thank David Rumelhart for useful discussions during the development of the basic architecture of TRACE and Eileen Conway, Mark Johnson, Dave Pare, and Paul Smith for their assistance in programing and graphics. Send requests for reprints to James L. McClelland, Department of Psychology, Carnegie–Mellon University, Schenley Park, Pittsburgh, PA 15213.

Consider the perception of the phoneme /g/ in the sentence "She received a valuable gift." There are a large number of cues in this sentence to the identity of this phoneme. First, there are the acoustic cues to the identity of the /g/ itself. Second, the other phonemes in the same word provide another source of cues, for if we know the rest of the phonemes in this word, there are only a few phonemes that can form a word with them. Third, the semantic and syntactic context further constrain the possible words which might occur, and thus limit still further the possible interpretation of the first phoneme in "gift."

There is ample evidence that all of these different sources of information are used in recognizing words and the phonemes they contain. Indeed, as Cole and Rudnicky (1983) have recently noted, these basic facts were described in early experiments by Bagley (1900) over 80 years ago. Cole and Rudnicky point out that recent work (which we consider in detail below) has added clarity and detail to these basic findings but has not lead to a theoretical synthesis that provides a satisfactory account of these and many other basic aspects of speech perception.

In this paper, we describe a model whose primary purpose is to account for the integration of multiple sources of information, or constraint, in speech perception. The model is constructed within a framework which appears to be ideal for the exploitation of simultaneous, and often mutual, constraints. This framework is the interactive activation framework (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1981, 1982). This approach grew out of a number of earlier ideas, some coming first from research on spoken language recognition (Marslen-Wilson & Welsh, 1978; Morton, 1969; Reddy, 1976) and others arising from more general considerations of interactive parallel processing (Anderson, 1977; Grossberg, 1978; McClelland, 1979).

According to the interactive-activation approach, information processing takes place through the excitatory and inhibitory interactions among a large number of processing elements called units. Each unit is a very simple processing device. It stands for a hypothesis about the input being processed. The activation of a unit is monotonically related