

## 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 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 (HNS-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 programming and graphics. Send requests for reprints to James L. McClelland, Department of Psychology, Carnegie-Mellon University, Schenley Park, Pittsburgh, PA 15213.

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.

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 Rudnicki (1983) have recently noted, these basic facts were described in early experiments by Bagley (1900) over 80 years ago. Cole and Rudnicki point out that recent work (which we consider in detail below) has added clarity and detail to these basic findings but has not led 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

to the strength of the hypothesis for which the unit stands. Constraints among hypotheses are represented by connections. Units which are mutually consistent are mutually excitatory, and units that are mutually inconsistent are mutually inhibitory. Thus, the unit for /g/ has mutually excitatory connections with units for words containing /g/, and has mutually inhibitory connections with units for other phonemes. When the activation of a unit exceeds some threshold activation value, it begins to influence the activation of other units via its outgoing connections; the strength of these signals depends on the degree of the sender's activation. The state of the system at a given point in time represents the current status of the various possible hypotheses about the input; information processing amounts to the evolution of that state, over time. Throughout the course of processing, each unit is continually receiving input from other units, continually updating its activation on the basis of these inputs, and, if it is over threshold, it is continually sending excitatory and inhibitory signals to other units. This "interactive-activation" process allows each hypothesis both to constrain and be constrained by other mutually consistent or inconsistent hypotheses.

#### *Criteria and Constraints on Model Development*

There are generally two kinds of models of the speech perception process. One kind of model, which grows out of speech engineering and artificial intelligence, attempts to provide a machine solution to the problem of speech recognition. Examples of this kind of model are HEARSAY (Erman & Lesser, 1980; Reddy, Erman, Fennell, & Neely, 1973) HWIM (Wolf & Woods, 1978), HARPY (Lowerre, 1976), and LAFS/SCRIBER (Klatt, 1980). A second kind of model, growing out of experimental psychology, attempts to account for aspects of psychological data on the perception of speech. Examples of this class of models include Marslen-Wilson's COHORT Model (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978; Nusbaum & Slowiaczek, 1982); Massaro's feature integration model (Massaro, 1981; Massaro & Oden, 1980a, 1980b; Oden & Massaro, 1978); Cole and Jakimik's (1978, 1980) model of auditory word processing, and the model of auditory and phonetic memory espoused by Fujisaki and Kawashima (1968) and Pisoni (1973, 1975).

Each approach honors a different criterion for success. Machine models are judged in terms of actual performance in recognizing real speech. Psychological models are judged in terms of their ability to account for details of human performance in speech recognition. We call these two criteria *computational* and *psychological* adequacy.

In extending the interactive activation approach to speech perception, we had essentially two questions: First, could the interactive-activation

approach contribute toward the development of a computationally sufficient framework for speech perception? Second, could it account for what is known about the psychology of speech perception? In short, we wanted to know, was the approach fruitful, both on computational and psychological grounds.

Two facts immediately became apparent. First, spoken language introduces many challenges that make it far from clear how well the interactive-activation approach will serve when extended from print to speech. Second, the approach itself is too broad to provide a concrete model, without further assumptions. Here we review several facts about speech that played a role in shaping the specific assumptions embodied in TRACE.

#### *Some Important Facts about Speech*

Our intention here is not to provide an extensive survey of the nature of speech and its perception, but rather to point to several fundamental aspects of speech that have played important roles in the development of the model we describe here. A very useful discussion of several of these points is available in Klatt (1980).

*Temporal nature of the speech stimulus.* It does not, of course, take a scientist to observe one fundamental difference between speech and print: speech is a signal which is extended in time, whereas print is a stimulus which is extended in space. The sequential nature of speech poses problems for a modeler, in that to account for context effects, one needs to keep a record of the context. It would be a simple matter to process speech if each successive portion of the speech input were processed independently of all of the others, but in fact, this is clearly not the case. The presence of context effects in speech perception requires a mechanism that keeps some record of that context, in a form that allows it to influence the interpretation of subsequent input.

A further point, and one that has been much neglected in certain models, is that it is not only prior context but also subsequent context that influences perception. (This and related points have recently been made by Grosjean & Gee, 1984; Salasoo & Pisoni, 1985; and Thompson, 1984). For example, Ganong (1980) reported that the identification of a syllable-initial speech sound that was constructed to be between /g/ and /k/ was influenced by whether the rest of the syllable was /is/ (as in "kiss") or /ift/ (as in "gift"). Such "right context effects" (Thompson, 1984) indicate that the perception of what comes in now both influences and is influenced by the perception of what comes in later. This fact suggests that the record of what has already been presented cannot not be a static representation, but should remain in a malleable form, subject to alteration as a result of influences arising from subsequent context.

*Lack of boundaries and temporal overlap.* A second fundamental point about speech is that the cues to successive units of speech frequently overlap in time. The problem is particularly severe at the phoneme level. A glance at a schematic speech spectrogram (Liberman, 1970; Fig. 1) clearly illustrates this problem. There are no separable packets of information in the spectrogram like the separate feature bundles that make up letters in printed words.

Because of the overlap of successive phonemes, it is difficult and, we believe, counterproductive to try to divide the speech stream up into separate phoneme units in advance of identifying the units. A number of other researchers (e.g., Fowler, 1984; Klatt, 1980) have made much the same point. A superior approach seems to be to allow the phoneme identification process to examine the speech stream for characteristic patterns, without first segmenting the stream into separate units.

The problem of overlap is less severe for words than for phonemes, but it does not go away completely. In rapid speech, words run into each other, and there are no pauses between words in running speech. To be sure, there are often cues that signal the locations of boundaries between words—stop consonants are generally aspirated at the beginnings of stressed words in English, and word initial vowels are generally preceded by glottal stops, for example. These cues have been studied by a number of investigators, particularly Lehiste (e.g., Lehiste, 1960, 1964) and Nakatani and collaborators. Nakatani and Dukas (1977) demonstrated that perceivers exploit some of these cues but found that certain utterances do not provide sufficient cues to word boundaries to permit reliable perception of the intended utterance. Speech errors often involve errors of

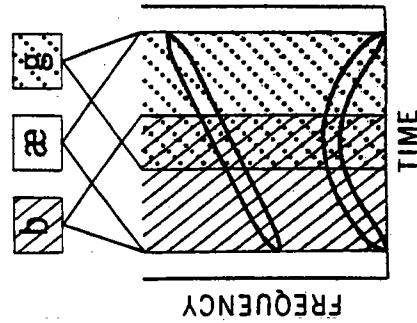


FIG. 1. A schematic spectrogram for the syllable "bag," indicating the overlap of the information specifying the different phonemes. Reprinted with permission from Liberman (1970).

word segmentation (Bond & Garnes, 1980), and certain segmentation decisions are easily influenced by contextual factors (Cole & Jakimik, 1980). Thus, it is clear that word recognition cannot count on an accurate segmentation of the phoneme stream into separate word units, and in many cases such a segmentation would perforce exclude from one of the words a shared segment that is doing double duty in each of two successive words.

*Context-sensitivity of cues.* A third major fact about speech is that the cues for a particular unit vary considerably with the context in which they occur. For example, the transition of the second formant carries a great deal of information about the identity of the stop consonant /b/ in Fig. 1, but that formant would look quite different had the syllable been "big" or "bog" instead of "bag." Thus the context in which a phoneme occurs restructures the cues to the identity of that phoneme (Liberman, 1970). The extent of the restructuring depends on the unit selected and on the particular cue involved. But the problem is ubiquitous in speech.

Not only are the cues for each phoneme dramatically affected by preceding and following context, they are also altered by more global factors such as rate of speech (Miller, 1981), by morphological and prosodic factors such as position in word and in the stress contour of the utterance, and by characteristics of the speaker such as size and shape of the vocal tract, fundamental frequency of the speaking voice, and dialectical variations (see Klatt, 1980, and Repp & Liberman, 1984, for discussions).

A number of different approaches to the problem have been tried by different investigators. One approach is to try to find relatively invariant—generally relational—features (e.g., Stevens & Blumstein, 1981). Another approach has been to redefine the unit so that it encompasses the context and therefore becomes more invariant (Fujimura & Lovins, 1982; Klatt, 1980; Wickelgren, 1969). While these are both sensible and useful approaches, the first has not yet succeeded in establishing a sufficiently invariant set of cues, and the second may alleviate but does not eliminate the problem; even units such as demissyllables (Fujimura & Lovins, 1982), context-sensitive allophones (Wickelgren, 1969), or even whole words (Klatt, 1980) are still influenced by context. We have chosen to focus instead on a third possibility: that the perceptual system uses information from the context in which an utterance occurs to alter connections, thereby effectively allowing the context to retune the perceptual mechanism on the fly.

*Noise and indeterminacy in the speech signal.* To compound all the problems alluded to above, there is the additional fact that speech is often perceived under less than ideal circumstances. While a slow and careful speaker in a quiet room may produce sufficient cues to allow correct

perception of all of the phonemes in an utterance without the aid of lexical or other higher level constraints, these conditions do not always obtain. People can correctly perceive speech under quite impoverished conditions, if it is semantically coherent and syntactically well formed (G. Miller, Heise, & Lichten, 1951). This means that the speech mechanisms must be able to function, even with a highly degraded stimulus. In particular, as Thompson (1984), Norris (1982), and Grosjean and Gee (1984) have pointed out, the mechanisms of speech perception cannot count on accurate information about any part of a word. As we shall see, this fact poses a serious problem for one of the best current psychological models of the process of spoken word recognition (Marslen-Wilson & Welsh, 1978).

Many of the characteristics that we have reviewed differentiate speech from print—at least, from very high quality print on white paper—but it would be a mistake to think that similar problems are not encountered in other domains. Certainly, the sequential nature of spoken input sets speech apart from vision, in which there can be some degree of simultaneity of perception. However, the problems of ill-defined boundaries, context sensitivity of cues, and noise and indeterminacy are central problems in vision just as much as they are in speech (cf. Ballard, Hinton, and Sejnowski, 1983; Barrow & Tenenbaum, 1978; Marr, 1982). Thus, though the model we present here is focussed on speech perception, we would hope that the ways in which it deals with the challenges posed by the speech signal are applicable in other domains.

### *The Importance of the Right Architecture*

All four of the considerations listed above played an important role in the formulation of the TRACE model. The model is an instance of an interactive activation model, but it is by no means the only instance of such a model that we have considered or that could be considered. Other formulations we considered simply did not appear to offer a satisfactory framework for dealing with these four aspects of speech (see Eelman & McClelland, 1984, for discussion). Thus, the TRACE model hinges as much on the particular processing architecture it proposes for speech perception as it does on the interactive activation processes that occur within this architecture.

Interactive-activation mechanisms are a class too broad to stand or fall on the merits of a single model. To the extent that computationally and psychologically adequate models can be built within the framework, the attractiveness of the framework as a whole is, of course, increased, but the adequacy of any particular model will generally depend on the particular assumptions that model embodies. It is no different with interactive-

activation models than with models in any other computational framework, such as expert systems or production systems.

## THE TRACE MODEL

### *Overview*

The TRACE model consists primarily of a very large number of units, organized into three levels, the *feature*, *phoneme*, and *word* levels. Each unit stands for a hypothesis about a particular perceptual object occurring at a particular point in time defined relative to the beginning of the utterance.

A small subset of the units in TRACE II, the version of the model we focus on in this paper, is illustrated in Figs. 2, 3, and 4. Each of the three figures replicates the same set of units, illustrating a different property of the model in each case. In the figures, each rectangle corresponds to a separate processing unit. The labels on the units and along the side indicate the spoken object (feature, phoneme, or word) for which each unit stands. The left and right edges of each rectangle indicate the portion of the input the unit spans.

At the feature level, there are several banks of feature detectors, one for each of several dimensions of speech sounds. Each bank is replicated for each of several successive moments in time, or time slices. At the phoneme level, there are detectors for each of the phonemes. There is one copy of each phoneme detector centered over every three time slices. Each unit spans six time slices, so units with adjacent centers span overlapping ranges of slices. At the word level, there are detectors for each word. There is one copy of each word detector centered over every three feature slices. Here each detector spans a stretch of feature slices corresponding to the entire length of the word. Again, then, units with adjacent centers span overlapping ranges of slices.

Input to the model, in the form of a pattern of activation to be applied to the units at the feature level, is presented sequentially to the feature-level units in successive slices, as it would if it were a real speech stream, unfolding in time. Mock-speech inputs on the three illustrated dimensions for the phrase "tea cup" (/tik p/) are shown in Fig. 2. At any instant, input is arriving only at the units in one slice at the feature level. In terms of the display in Fig. 2, then, we can visualize the input being applied to successive slices of the network at successive moments in time. However, it is important to remember that all the units are continually involved in processing, and processing of the input arriving at one time is just beginning as the input is moved along to the next time slice.

The entire network of units is called "the Trace," because the pattern of activation left by a spoken input is a trace of the analysis of the input at each of the three processing levels. This trace is unlike many traces,

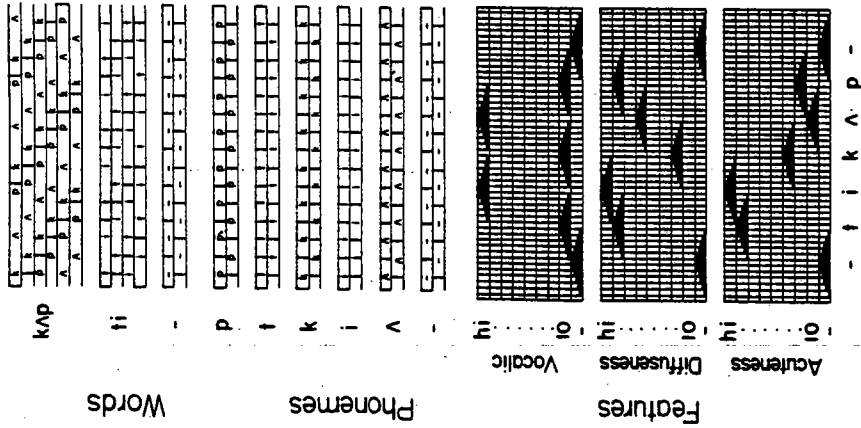


FIG. 2. A subset of the units in TRACE II. Each rectangle represents a different unit. The labels indicate the item for which the unit stands, and the horizontal edges of the rectangle indicate the portion of the Trace spanned by each unit. The input feature specifications for the phrase "tea cup," preceded and followed by silence, are indicated for the three illustrated dimensions by the blackening of the corresponding feature units.

though, in that it is dynamic, since it consists of activations of processing elements, and these processing elements continue to interact as time goes on. The distinction between perception and (primary) memory is completely blurred; since the percept is unfolding in the same structures that serve as working memory, and perceptual processing of older portions of the input continues even as newer portions are coming into the system. These continuing interactions permit the model to incorporate right context effects, and allow the model to account directly for certain aspects

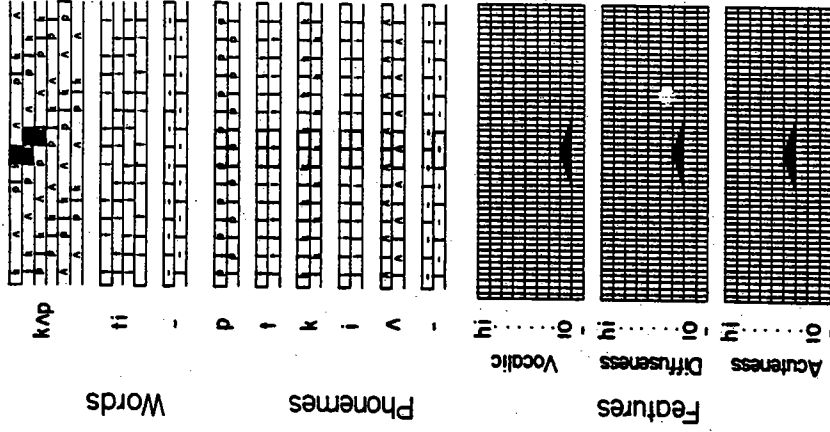


FIG. 3. The connections of the unit for the phoneme /k/, centered over Time Slice 24. The rectangle for this unit is highlighted with a bold outline. The /k/ unit has mutually excitatory connections to all the word- and feature-level units colored either partly or wholly in black. The more coloring on a units' rectangle, the greater the strength of the connection. The /k/ unit has mutually inhibitory connections to all of the phoneme-level units colored partly or wholly in grey. Again, the relative amount of inhibition is indicated by the extent of the coloring of the unit; it is directly proportional to the extent of the temporal overlap of the units.

of short-term memory, such as the fact that more information can be retained for short periods of time if it hangs together to form a coherent whole.

Processing takes place through the excitatory and inhibitory interactions of the units in the Trace. Units on different levels that are mutually consistent have mutually excitatory connections, while units on the same

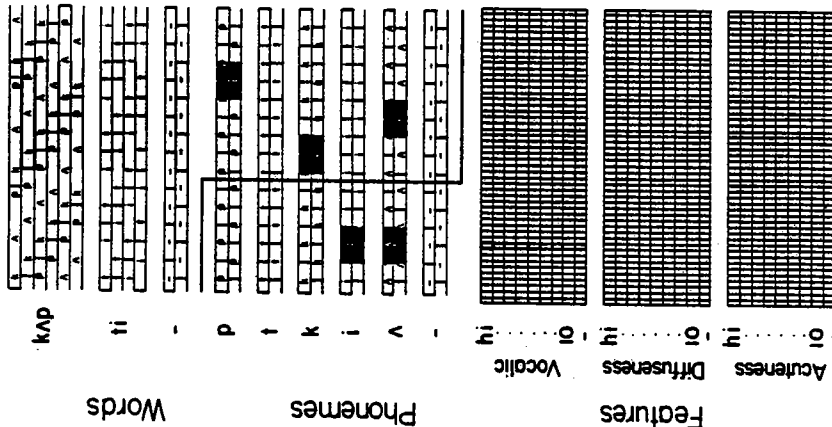


FIG. 4. The connections of the highlighted unit for the high value on the Vocalic feature dimension in Time Slice 9 and for the highlighted unit for the word /k p/ starting in Slice 24. Excitatory connections are represented in black, inhibitory connections in grey, as in Fig. 3.

level that are inconsistent have mutually inhibitory connections. All connections are bidirectional. Bidirectional excitatory and inhibitory connections of the unit for /k/ centered over Feature-slice 24 (counting from 0) are shown in Fig. 3; connections for the high value of the feature Vocalic in Slice 9 and for the word /k p/ with the /k/ centered over Slice 24 are shown in Fig. 4.

The interactive activation model of visual word recognition (McClelland & Rumelhart, 1981) included inhibitory connections between each unit on the feature level and letters that did not contain the feature, and between each letter unit and the words that did not contain the letter. Thus the units for *T* in the first letter position inhibited the units for all words that did not begin with *T*. However, more recent versions of the

visual model eliminate these between-level inhibitory connections, since these connections can interfere with successful use of partial information (McClelland, 1985; McClelland, 1986). Like these newer versions of the visual model, TRACE likewise contains no between-level inhibition. We will see that this feature of TRACE plays a very important role in its ability to simulate a number of empirical phenomena.

*Sources of TRACE's architecture.* The inspiration for the architecture of TRACE goes back to the HEARSAY Speech understanding system (Erman & Lesser, 1980; Reddy et al., 1973). HEARSAY introduced the notion of a Blackboard, a structure similar to the Trace in the TRACE model. The main difference is that the Trace is a dynamic processing structure that is self-updating, while the Blackboard in HEARSAY was a passive data structure through which autonomous processes shared information.

The architecture of TRACE bears a strong resemblance to the "neural spectrogram" proposed by Crowder (1978, 1981) to account for interference effects between successive items in short-term memory. Like our Trace, Crowder's neural spectrogram provides a dynamic working memory representation of a spoken input. There are two important differences between the Trace and Crowder's neural spectrogram, however. First of all, the neural spectrogram was assumed only to represent the frequency spectrum of the speech wave over time; the Trace, on the other hand, represents the speech wave in terms of a large number of different feature dimensions, as well as in terms of the phonemes and words consistent with the pattern of activation at the feature level. In this regard TRACE might be seen as an extension of the neural spectrogram idea. The second difference is that Crowder postulates inhibitory interactions between detectors for spectral components spaced up to several hundred milliseconds apart. These inhibitory interactions extend considerably farther than those we have included in the feature level of the Trace. This difference does not reflect a disagreement with Crowder's assumptions. Though we have not found it necessary to adopt this assumption to account for the phenomena we focus on in this article, lateral extension of inhibition in the time domain might well allow the TRACE framework to incorporate many of the findings Crowder discusses in the two articles cited.

#### *Context-Sensitive Tuning of Phoneme Units*

The connections between the feature and phoneme level determine what pattern of activations over the feature units will most strongly activate the detector for each phoneme. To cope with the fact that the features representing each phoneme vary according to the phonemes surrounding them, the model adjusts the connections from units at the feature level to units at the phoneme level as a function of activations at the

phoneme level in preceding and following time slices. For example, when the phoneme /t/ is preceded or followed by the vowel /i/, the feature pattern corresponding to the /t/ is very different than it is when the /t/ is preceded or followed by another vowel, such as /a/. Accordingly, when the unit for /i/ in a particular slice is active, it changes the pattern of connections for units for /t/ in preceding and following slices.

#### TRACE I and TRACE II

In developing TRACE, and in trying to test its computational and psychological adequacy, we found that we were sometimes led in rather different directions. We wanted to show that TRACE could process real speech, but to build a model that did so it was necessary to worry about exactly what features must be extracted from the speech signal, about differences in duration of different features of different phonemes, and about how to cope with the ways in which features and feature durations vary as a function of context. Obviously, these are important problems, worthy of considerable attention. However, concern with these issues tended to obscure attention to the fundamental properties of the model and the model's ability to account for basic aspects of the psychological data obtained in many experiments.

To cope with these conflicting goals, we have developed two different versions of the model, called TRACE I and TRACE II. Both models spring from the same basic assumptions, but focus on different aspects of speech perception. TRACE I was designed to address some of the challenges posed by the task of recognizing phonemes from real speech. This version of the model is described in detail in Elman and McClelland (in press). With this version of the model, we were able to show that the TRACE framework could indeed be used to process real speech—albeit from a single speaker uttering isolated monosyllables at this point. We were also able to demonstrate the efficacy of the idea of adjusting feature to phoneme connections on the basis of activations produced by surrounding context. With connection strength adjustment in place, the model was able to identify the stop consonant in 90% of a set of isolated monosyllables correctly, up from 79% with an invariant set of connections. This level of performance is comparable to what has been achieved by other machine-based phoneme identification schemes (e.g., Kopec, 1984) and illustrates the promise of the connection strength adjustment scheme for coping with variability due to local phonetic context. Ideas for extending the connection strength adjustment scheme to deal with the ways in which cues to phoneme identification vary with global variables (rate, speaker characteristics, etc.) are considered in the general discussion.

TRACE II, the version described in the present paper, was designed to account primarily for lexical influences on phoneme perception and

for what is known about on-line recognition of words, though we use it to illustrate how certain other aspects of phoneme perception fall out of the TRACE framework. This version of the model is actually a simplified version of TRACE I. Most importantly, we eliminated the connection-strength adjustment facility, and we replaced the real speech inputs to TRACE I with mock speech. This mock speech input consisted of overlapping but contextually invariant specifications of the features of successive phonemes. Obviously, then, TRACE II sidesteps many fundamental issues about speech. But it makes it much easier to see how the mechanism can account for a number of aspects of phoneme and word recognition. A number of further simplifying assumptions were made to facilitate examination of basic properties of the interactive activation processes taking place within the model.

The following sections describe TRACE II in more detail. First we consider the specifications of the mock-speech input to the model, and then we consider the units and connections that make up the Trace at each of the three levels.

#### Mock-Speech Inputs

The input to TRACE II was a series of specifications for inputs to units at the feature level, one for each 25-ms time slice of the mock utterance. These specifications were generated by a simple computer program from a sequence of to-be-presented segments provided by the human user of the simulation program. The allowed segments consisted of the stop consonants /b/, /p/, /d/, /t/, /g/, and /k/, the fricatives /s/ and /ʃ/ ("sh" as in "ship"), the liquids /l/ and /r/, and the vowels /a/ (as in "pot"), /i/ (as in "beet"), /u/ (as in "boot"), and /-/ (as in "but"). /-/ was also used to represent reduced vowels such as the second vowel in "target." There was also a "silence" segment represented by /-. Special segments, such as a segment halfway between /b/ and /p/, were also used; their properties are described in descriptions of the relevant simulations.

A set of seven dimensions was used in TRACE II to represent the feature-level inputs. Five of the dimensions (Consonantal, Vocalic, Dif-fuseness, Acuteness, and Voicing) were taken from classical work in phonology (Jakobson, Fant, & Halle, 1952), though we treat each of these dimensions as continua, in the spirit of Oden and Massaro (1978), rather than as binary features. A sixth dimension, Power, was included because it has been found useful for phoneme identification in various machine systems (e.g., Reddy, 1976), and it was incorporated here to add an additional dimension to increase the differentiation of the vowels and consonants. The seventh dimension, the amplitude of the burst of noise that occurs at the beginning of word initial stops, was included to provide an additional basis for distinguishing the stop consonants, which otherwise differed from each other on only one or two dimensions. Of course, these

dimensions are intentional simplifications of the real acoustic structure of speech, in much the same way that the font used by McClelland and Rumelhart (1981) in the interactive-activation model of visual word recognition was an intentional simplification of the real structure of print.

Each dimension was divided into eight value ranges. Each phoneme was assigned a value on each dimension; the values on the Vocalic, Diffuseness, and Acuteness dimensions for the phonemes in the utterance /tik'p/ are shown in Fig. 2. The full set of values are shown in Table 1. Numbers in the cells of the table indicate which value on the indicated dimension was most strongly activated by the feature pattern for the indicated phoneme. Values range from 1 = *very low* to 8 = *very high*. The last two dimensions were altered for the categorical perception and trading relations simulations.

Values were assigned to approximate the values real phonemes would have on these dimensions and to make phonemes that fall into the same phonetic category have identical values on many of the dimensions. Thus, for example, all stop consonants were assigned the same values on the Power, Vocalic, and Consonantal dimensions. We do not claim to have captured the details of phoneme similarity exactly. Indeed, one cannot do so in a fixed feature set because the similarities vary as a function of context. However, the feature sets do have the property that the feature pattern for one phoneme is more similar to the feature pattern for other phonemes in the same phonetic category (stop, fricative, liquid, or vowel) than it is to the patterns for phonemes in other categories. Among the stops, those phonemes sharing place of articulation or voicing are more similar than those sharing neither attribute.

The correlations of the feature patterns for the 15 phonemes used are shown in Table 2. It is these correlations of the patterns assigned to the

TABLE 1  
Phoneme Feature Values Used in TRACE II

Phoneme	Power	Vocalic	Diffuse	Acute	Cons.	Voiced	Burst
p	4	1	7	2	8	1	8
b	4	1	7	2	8	7	7
t	4	1	7	7	8	1	6
d	4	1	7	7	8	7	5
k	4	1	2	3	8	1	4
g	4	1	2	3	8	7	3
s	6	4	7	8	5	1	—
z	6	4	6	4	5	1	—
r	7	7	1	2	3	8	—
l	7	7	2	4	3	8	—
a	8	8	2	1	1	8	—
i	8	8	8	8	1	8	—
u	8	8	6	2	1	8	—
·	7	8	5	1	1	8	—

TABLE 2  
Correlations of Feature Patterns of the Different Phonemes Used in TRACE II

Phoneme	p	b	t	d	k	g	s	z	r	l	a	i	u
p	—	.76	.71	.56	.46	.60	.30	.46	.35	.24	.65	.20	.37
b	.76	—	.56	.71	.46	.60	.30	.46	.35	.24	.65	.20	.37
t	.71	.56	—	.76	.42	.56	.35	.42	.77	.77	.24	.65	.37
d	.56	.71	.76	—	.42	.56	.35	.42	.77	.77	.24	.65	.37
k	.46	.46	.42	.56	—	.60	.30	.42	.77	.77	.24	.65	.37
g	.60	.60	.56	.42	.60	—	.30	.42	.77	.77	.24	.65	.37
s	.30	.30	.35	.35	.35	.30	—	.30	.35	.35	.24	.65	.37
z	.46	.46	.42	.42	.42	.46	.30	—	.42	.42	.24	.65	.37
r	.60	.60	.56	.56	.56	.60	.30	.46	—	.56	.24	.65	.37
l	.60	.60	.56	.56	.56	.60	.30	.46	.56	—	.24	.65	.37
a	.76	.76	.71	.71	.71	.76	.30	.76	.71	.71	—	.65	.32
i	.76	.76	.71	.71	.71	.76	.30	.76	.71	.71	.65	—	.32
u	.76	.76	.71	.71	.71	.76	.30	.76	.71	.71	.65	.32	—

Note. Correlations of less than .20 have been replaced by blanks.

different phonemes, rather than the actual values assigned to particular phonemes or even the labels attached to the different mock-speech dimensions, that determine the behavior of the simulation model, since it is these correlations that determine how much an instance of one phoneme will tend to excite the detector for another.

The feature patterns were constructed in such a way that it was possible to create feature patterns that would activate two different phonemes in the same category (stop, liquid, fricative, or vowel) to an equal extent by averaging the values of the two phonemes on one or more dimensions. In this way, it was a simple matter to make up ambiguous inputs, halfway between two phonemes, or to construct continua varying between two phonemes on one or more dimensions.

The feature specification of each phoneme in the input stream extended over 11 time slices of the input. The strength of the pattern grew to a peak at the 6th slice and fell off again, as illustrated in Fig. 2. Peaks of successive phonemes were separated by 6 slices. Thus, specifications of successive phonemes overlapped, as they do in real speech (Fowler, 1984; Liberman, 1970).

Generally, there were no cues to word boundaries in the speech stream—the feature specification for the last phoneme of one word overlapped with the first phoneme of the next in just the same way feature specifications of adjacent phonemes overlap within words. However, entire utterances presented to the model for processing—whether they were individual syllables, words, or strings of words—were preceded and followed by silence. Silence was not simply the absence of any input; rather, it was a pattern of feature values, just like the phonemes. Thus, a ninth value on each of the seven dimensions was associated with silence. These values were actually outside the range of values which occurred in the phonemes themselves, so that the features of silence were completely uncorrelated with the features of any of the phonemes used.

#### *Feature Level Units and Connections*

The units at the feature level are detectors for features of the speech stream at particular moments in time. In TRACE II, there was a unit for each of the nine values on each of the seven dimensions in each time slice of the Trace. The figures show three sets of feature units in several time slices. Units for features on the same dimension within the same time slice are mutually inhibitory. Thus, the unit for the high value of the vocalic dimension in Time Slice 9 inhibits the units for other values on the same dimension in the same time slice, as illustrated in Fig. 4. This figure also illustrates the mutually excitatory connections of this same feature unit with units at the phoneme level. In the next section we describe these connections from the point of view of the phoneme level.

#### *The Phoneme Level and Feature-Phoneme Connections*

At the phoneme level, there is a set of detectors for each of the 15 phonemes listed above. In addition, there is a set of detectors for the presence of silence. These silence detectors are treated like all other phoneme detectors. Each member of the set of detectors for a particular phoneme is centered over a different time slice at the feature level, and the centers are spaced three time slices apart. The unit centered over a particular slice received excitatory input from feature units in a range of slices, extending both forward and backward from the slice in which the phoneme unit is located. It also sends excitatory feedback down to the same feature units in the same range of slices.

The connection strengths between the feature-level units and a particular phoneme-level unit exactly match the feature pattern the phoneme is given in its input specification. Thus, as illustrated in Fig. 3, the strengths of the connections between the node for /k/ centered over Time Slice 24 and the nodes at the feature level are exactly proportional to the pattern of input to the feature level produced by an input specification containing the features of /k/ centered in the same time slice.

There are inhibitory connections between units at the phoneme level. Units inhibit each other to the extent that the speech objects they stand for represent alternative interpretations of the content of the speech stream at the same point in the utterance. Note that, although the feature specification of a phoneme is spread over a window of 11 slices, successive phonemes in the input have their centers 6 slices apart. Thus each phoneme-level unit is thought of as spanning 6 feature-level slices, as illustrated in Fig. 3. Each unit inhibits others in proportion to their overlap. Thus, a phoneme detector inhibits other phoneme detectors centered over the same slice twice as much as it inhibits detectors centered 3 slices away, and inhibits detectors centered 6 or more slices away not at all.

#### *Word Units and Word-Phoneme Connections*

There is a unit for every word in every time slice. Each of these units represents a different hypothesis about a word identity and starting location in the Trace. For example, the unit for the word /k'p/ in Slice 24 (highlighted in Fig. 4) represents the hypothesis that the input contains the word "cup" starting in Slice 24. More exactly, it represents the hypothesis that the input contains the word "cup" with its first phoneme centered in Time Slice 24.

Word units receive excitation from the units for the phonemes they contain in a series of overlapping windows. Thus, the unit for "cup" in Time Slice 24 will receive excitation from /k/ in slices neighboring Slice

24, from /r/ in slices neighboring Slice 30, and from /p/ in slices neighboring Slice 36. As with the feature-phoneme connections, these connections are strongest at the center of the window and fall off linearly on either side.

The inhibitory connections at the word level are similar to those at the phoneme level. Again, the strength of the inhibition between two word units depends on the number of time slices in which they overlap. Thus, units representing alternative interpretations of the same stretch of phoneme units are strongly competitive, but units representing interpretations of nonoverlapping sequences of phonemes do not compete at all.

TRACE II has detectors for the 211 words found in a computerized phonetic word list that met all of the following constraints: (a) the word consisted only of the phonemes listed above; (b) it was not an inflection of some other word that could be made by adding "-ed," "-s," or "-ing"; (c) the word together with its "-ed," "-s," and "-ing" inflections occurred with a frequency of 20 or more per million in the Kucera and Francis (1967) word count. It is not claimed that the model's lexicon is an exhaustive list of words meeting this criterion, since the computerized phonetic lexicon was not complete, but it is reasonably close to this. To make specific points about the behavior of the model, detectors for the following three words not in the main list were added: "blush," "regal," and "sleet." The model also had detectors at the word level for silence (-/), which was treated like a one-phoneme word.

#### *Presentation and Processing of an Utterance*

Before processing of an utterance begins, the activations of all of the units are set at their resting values. At the start of processing, the input to the initial slice of feature units is applied. Activations are then updated, ending the initial time cycle. On the next time cycle, the input to the next slice of feature units is applied, and excitatory and inhibitory inputs to each unit resulting from the pattern of activation left at the end of the previous time slice are computed.

It is important to remember that the input is applied, one slice at a time, proceeding from left to right as though it were an ongoing stream of speech "writing on" the successive time slices of the Trace. The interactive-activation process is occurring throughout the Trace on each time slice, even though the external bottom-up input is only coming into the feature units one slice at a time. Processing interactions can continue even after the left to right sweep through the input reaches the end of the Trace. Once this happens, there are simply no new input specifications applied to the Trace; the continuing interactions are based on what has already been presented. This interaction process is assumed to continue

indefinitely, though for practical purposes it is always terminated after some predetermined number of time cycles has elapsed.

#### *Details of Processing Dynamics*

The interactive activation process in the Trace model follows the dynamic assumptions laid out in McClelland and Rumelhart (1981). Each unit has a resting activation value arbitrarily set at 0, a maximum activation value arbitrarily set at 1.0, and a minimum activation set at  $-0.3$ . On every time cycle of processing, all the weighted excitatory and inhibitory signals impinging upon a unit are added together. The signal from one unit to another is just the extent to which its activation exceeds 0; if its activation is less than 0, the signal is 0.<sup>1</sup> Global level-specific excitatory, inhibitory, and decay parameters scale the relative magnitudes of different types of influences on the activation of each unit. Values for these parameters are given below.

After the net input to each unit has been determined based on the prior activations of the units, the activations of the units are all updated for the next processing cycle. The new value of the activation of the unit is a function of its net input from other units and its previous activation value. The exact function used (see McClelland & Rumelhart, 1981) keeps unit activations bounded between their maximum and minimum values. Given a constant input, the activation of a unit will stabilize at a point between its maximum and minimum that depends on the strength and sign (excitatory or inhibitory) of the input. With a net input of 0, the activation of the unit will gradually return to its resting level.

Each processing time cycle corresponds to a single time slice at the feature level. This is actually a parameter of the model—there is no intrinsic reason why there should be a single cycle of the interactive-activation process synchronized with the arrival of each successive slice of the input. A higher rate of cycling would speed the percolation of effects of new input through the network relative to the rate of presentation.

#### *Output Assumptions*

Activations of units in the Trace rise and fall as the input sweeps across the feature level. At any time, a decision can be made based on the pattern of activation as it stands at that moment. The decision mechanism can, we assume, be directed to consider the set of units located within a small window of adjacent slices within any level. The units in this set then

<sup>1</sup> At the word level, the inhibitory signal from one word to another is just the square of the extent to which the sender's activation exceeds zero. This tends to smooth the effects of many units suddenly becoming slightly activated, and of course it also increases the dominance of one active word over many weakly activated ones.

constitute the set of response alternatives, designated by the identity of the item for which the unit stands (note that with several adjacent slices included in the set, several units in the alternative set may correspond to the same overt response). Word identification responses are assumed to be based on readout from the word level, and phoneme identification responses are assumed to be based on readout from the phoneme level. As far as phoneme identification is concerned, then, a homogeneous mechanism is assumed to be used with both word and nonword stimuli. The decision mechanism can be asked to make a response either (a) at a critical time during processing, or (b) when a unit in the alternative set reaches a critical strength relative to the activation of other alternative units. Once a decision has been made to make a response, one of the alternatives is chosen from the members of the set. The probability of choosing a particular alternative  $i$  is then given by the Luce (1959) choice rule:

$$p(R_i) = \frac{S_i}{\sum_j S_j}$$

when  $j$  indexes the members of the alternative set, and

$$S_j = e^{kx_j}$$

The exponential transformation ensures that all activations are positive and gives great weight to stronger activations, and the Luce rule ensures that the sum of all of the response probabilities adds up to 1.0. Substantially the same assumptions were used by McClelland and Rumelhart (1981).

#### Minimizing the Number of Parameters

At the expense of considerable realism, we have tried to keep TRACE II simple by using homogeneous parameters wherever possible. Thus, as already noted, the feature specifications of all phonemes were spread out over the same number of time slices, effectively giving all phonemes the same duration. The strength of the total excitation coming into a particular phoneme unit from the feature units was normalized to the same value for all phonemes, thus making each phoneme equally excitable by its own canonical pattern. Other simplifying assumptions should be noted as well. For example, there were no differences in connections or resting levels for words of different frequency. It would have been a simple matter to incorporate frequency as McClelland and Rumelhart (1981) did, and a complete model would, of course, include some account for the ubiquitous effects of word frequency. We left it out here to facilitate an examination of the many other factors that appear to influence the process of word recognition in speech perception.

Even with all the simplifications described above, the TRACE model still has a number of free parameters. These parameters are listed in Table 3. It should be noted that parameters are not in general directly comparable across levels. For example, phoneme-to-phoneme and word-to-word inhibition are not directly comparable to each other or to feature-to-phoneme inhibition, since feature-level units compete only within a single slice, while phoneme and word units compete in proportion to their overlap.

There was some trial and error in finding the set of parameters used in the reported simulations, but, in general, the qualitative behavior of the model was remarkably robust under parameter variations, and no systematic search of the space of parameters was necessary. Generally, manipulations of parameters simply influence the magnitude or the timing of one effect or another without changing the basic nature of the effects observed. For example, stronger bottom-up excitation speeds things up and can indirectly influence the size of top-down effects, since, for example, stronger word level activations produce stronger feedback to the phoneme level. Stronger top-down excitation, of course, directly influences the magnitude of lexical effects. The one parameter that appeared to influence the qualitative behavior of the model was the strength of within-level inhibition. Stronger within-level inhibition make the model commit itself more strongly to slight early differences in activation among competing alternatives. There was, therefore, some tuning of this parameter to avoid early overcommitment that would prevent right context from exerting an influence under some circumstances. Finally, a low rate of feature-level decay was used to allow feature-level activations to persist after the input moved on to later slices.

The parameter values were held constant at the values shown in the

TABLE 3  
Parameters of TRACE II

Parameter	Value
Feature-phoneme excitation	.02
Phoneme-word excitation	.05
Word-phoneme excitation	.03
Phoneme-feature excitation	.00
Feature-level inhibition	.04
Phoneme-level inhibition <sup>a</sup>	.03
Word-level inhibition <sup>a</sup>	.04
Feature-level decay	.01
Phoneme-level decay	.03
Word-level decay	.05

<sup>a</sup> Per three time-slices of overlap.

table throughout the simulations, except in the simulations of categorical perception and trading relations. Since we were not explicitly concerned with the effects of feedback to the feature level in any of the other simulations, we set the feedback from the phoneme level to the feature level to zero to speed up the simulations in all other cases. In the categorical perception and trading relations simulations this parameter was set at .05. Phoneme-to-feature feedback tended to slow the effective rate of decay at the feature level and to increase the effective distinctiveness of different feature patterns. Rate of decay of feature-level activations and strength of phoneme-to-phoneme competition were set to .03 and .05 to compensate for these effects. No lexicon was used in the categorical perception and trading relations simulations, which is equivalent to setting the phoneme to word excitation parameter to zero.

#### THE DYNAMICS OF PHONEME PERCEPTION

In the introduction, we motivated the approach taken in the TRACE model in general terms. In this section, we see that the simple concepts that lead to TRACE provide a coherent and synthetic account of a large number of different kinds of findings on the perception of phonemes. Previous models have been able to provide fairly accurate accounts of a number of these phenomena. For example, Massaro and Oden's feature integration model (Massaro, 1981; Massaro & Oden, 1980a, 1980b; Oden & Massaro, 1978) accounts in detail for a large body of data on the influences of multiple cues to phoneme identity, and the Pisoni/Fujisaki-Kawashima model of categorical perception (Fujisaki & Kawashima, 1968; Pisoni, 1973, 1975) accounts for a large body of data on the conditions under which subjects can discriminate sounds within the same phonetic category. Marsler-Wilson's COHORT model can account for the time course of lexical influences on phoneme identification. What we hope to show here is that TRACE brings these phenomena, and several others not considered by either model, together into a coherent picture of the process of phoneme perception as it unfolds in time.

The present section consists of three main parts. The first focuses on lexical effects on phoneme identification and the conditions under which these effects are obtained. Here, we see how TRACE can account for the basic lexical effect, and we make it clear why lexical effects are only obtained under some conditions. The second part of this section focuses on the question of the role of phonotactic rules—that is, rules specifying which phonemes can occur together in English—in phoneme identification. Here, we see how TRACE mimics the apparently rule-governed behavior of human subjects, in terms of a "conspiracy" of the lexical items that instantiate the rule. The third part focuses on two aspects of phoneme identification often considered quite separately from lexical ef-

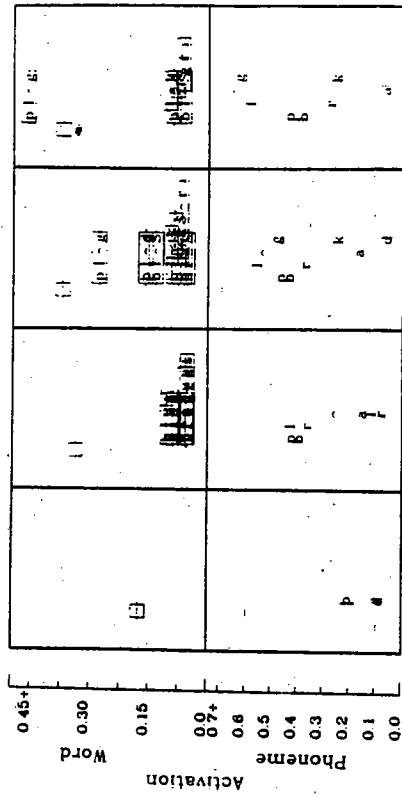


Fig. 5. Phoneme- and word-level activations at several points in the unfolding of a segment ambiguous between /b/ and /p/, followed by /l/, /i/, and /g/. See text for a full explanation.

fects—namely, the contrasting phenomena of cue tradeoffs in phoneme perception and categorical perception. Here we see that TRACE provides an account of both effects as well as details of their time course. All three parts of this section illustrate how the simple mechanisms of mutual excitation and inhibition among the processing units of the Trace provide a natural way of accounting for the relevant phenomena. The section ends with a brief consideration of the ways in which TRACE might be extended to cope with several other aspects of phoneme identification and perception.

#### Lexical Effects

*You can tell a phoneme by the company that it keeps.*<sup>2</sup> In this section, we describe a simple simulation of the basic lexical effect on phoneme identification reported by Ganong (1980). We start with this phenomenon because it, and the related phonemic restoration effect, were among the primary reasons why we felt that the interactive-activation approach would be appropriate for speech perception as well as visual word recognition and reading.

For the first simulation, the input to the model consisted of a feature specification which activated /b/ and /p/ equally, followed by (and partially overlapping with) the feature specifications for /l/, then /i/, then /g/. Figure 5 shows phoneme and word-level activations at several points in the unfolding of this input specification. Each panel of the figure represents

<sup>2</sup> This title is adapted from the title of a talk by David E. Rumelhart on related phenomena in letter perception. These findings are described in Rumelhart and McClelland (1982). We thank Dave for his permission to adapt the title.

a different point in time during the presentation and concomitant processing of the input. The upper portion of each panel is used to display activations at the word level; the lower panel is used for activations at the phoneme level. Each unit is represented by a rectangle, labeled with the identity of the item the unit stands for. The horizontal extension of the rectangle indicates the portion of the input spanned by the unit. The vertical position of the rectangle indicates the degree of activation of the unit. In this and subsequent figures, activations of the phoneme units located between the peaks of the input specifications of the phonemes (at Slices 3, 9, 15, etc.) have been deleted from the display for clarity (the activations of these units generally get suppressed by the model, since the units on the peaks tend to dominate them). The input itself is indicated below each panel, with the successive phonemes positioned at the temporal positions of the centers of their input specifications. The  $/r/$  along the x axis represents the point in the presentation of the input stream at which the snapshot was taken.

The figure illustrates the gradual buildup of activation of the two interpretations of the first phoneme, followed by gradual buildups in activation for subsequent phonemes. As these processes unfold, they begin to produce word-level activations. It is difficult to resolve any word-level activations in the first few frames, however, since in these frames, the information at the phoneme level simply has not evolved to the point where it provides enough constraint to select any one particular word. In this case, it is only after the  $/g/$  has come in that the model has information telling it whether the input is closer to "plug," "plus," "blush," or "blood" (TRACE's lexicon contains no other words beginning with  $/pl/$  or  $/bl/$ ). After that point, as illustrated in the fourth panel, "plug" wins the competition at the word level and, through feedback support to  $/p/$ , causes  $/p/$  to dominate  $/b/$  at the phoneme level. The model, then, provides an explicit account for the way in which lexical information can influence phoneme identification.

Two things about the lexical effect observed in this case are worthy of note. First, the effect is rather small. Second, it does not emerge until well after the ambiguous segment itself has come and gone. There is a slight advantage of  $/p/$  over  $/b/$  in Frames 2 and 3 of the figure. In these cases, however, the advantage is not due to the specific information that this item is the word "plug"—the model can have no way of knowing this at these points in processing. The slight advantage for  $/p/$  at these early points is due to the fact that there are more words beginning with  $/p/$  than  $/b/$  in the model's lexicon, and in particular, there are more beginning with  $/pl/$  than  $/bl/$ . So, when the input is  $/rld/$ , with the ? standing for the ambiguous  $/b/-p/$  segment, the model must actually overcome this slight  $/p/-ward$  bias. Eventually, it does so.

Figure 6 shows the temporal course of the strength of the

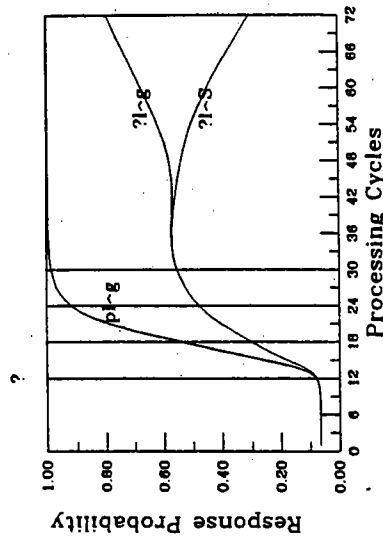


FIG. 6. The time course of the buildup in the strength of the  $/p/$  response based on activations of phoneme units in Slice 12, in processing an ambiguous  $/b/-p/$  segment in  $/rld/$ , and the same segment in  $/r-l's/$ . The ambiguous segment is indicated by the "?". Also shown is the buildup of response strength for processing an unambiguous  $/p/$  segment in  $/pl'g/$ . The vertical line topped with "?" indicates the point in time corresponding to the center of the initial segment in the input stream. Successive vertical lines indicate centers of successive phonemes.

$/p/$  response based on activations of the phoneme units in Slice 12 for two cases in which the initial segment is ambiguous between  $/p/$  and  $/b/$ . In one case, the ambiguous segment is followed by  $/l'g/$  (as in "plug"); in the other, it is followed by  $/l's/$  (as in "blush"). Given the model's restricted lexicon, which does not contain the word "plush," the lexical effect should lead to eventual dominance of the  $/p/$  response in the first case, but a suppression of the  $/p/$  response in the second case. The differences between the contexts do not begin to show up until after the center of the final phoneme, which occurs at Slice 30. The reason for this is simply that the information is not available until that point, because the phoneme that signals what the word will be comes at the very end of the word. The effect takes another few time slices to begin to influence the activation of the initial phoneme, because it percolates to the first phoneme by way of the feedback from the word or words that contain it.

*Elimination of the lexical effect by time pressure.* Fox (1982) has reported that the lexical effect on word initial segments is eliminated if subjects are given a deadline to respond within 500 ms of the ambiguous segment. Though they can correctly identify unambiguous segments in responses made before the deadline, these early responses show no sensitivity to the lexical status of the alternatives. Similar findings are also reported by Fox (1984).

Our model is completely consistent with Fox's results. Indeed, we have

already seen that the activations in the Trace only begin to reflect the lexical effect about one phoneme or so after the phoneme that establishes the lexical identity of the item. Given that this segment does not occur in Fox's experiments, until the second or third segment after the ambiguous segment, there is no way that a lexical effect could be observed in early responses.

But what about the fact that early responses to unambiguous segments can be accurate? TRACE accounts for this too. In Figure 7 we show the state of the Trace at various different points after the unambiguous /b/ in /b'g/. Here, the /b/ dominates the /p/ from the earliest point. The analogous result is obtained, when the stimulus is /p/ in /p'g/, and the activation for the initial phoneme is quite independent of whether or not the item is a word. The response strength for the case when /p'g/ is presented in Fig. 6 shows that the probability of choosing /p/ is near unity within 12 processing cycles, or 300 ms of the initial segment, well before the deadline would be reached—and well before word identity specifying information is available.

*Lexical effects late in a word.* In the model, lexical effects on word-initial segments develop rather late, at least in the case where there is no context preceding the word. Of course, the exact timing of the development of any lexical effect would be dependent upon the set of words activated by the stimulus; if one word predominated early on, a lexical effect could develop rather earlier. In general, though, word-initial ambiguities will require time to resolve on the basis of lexical information.

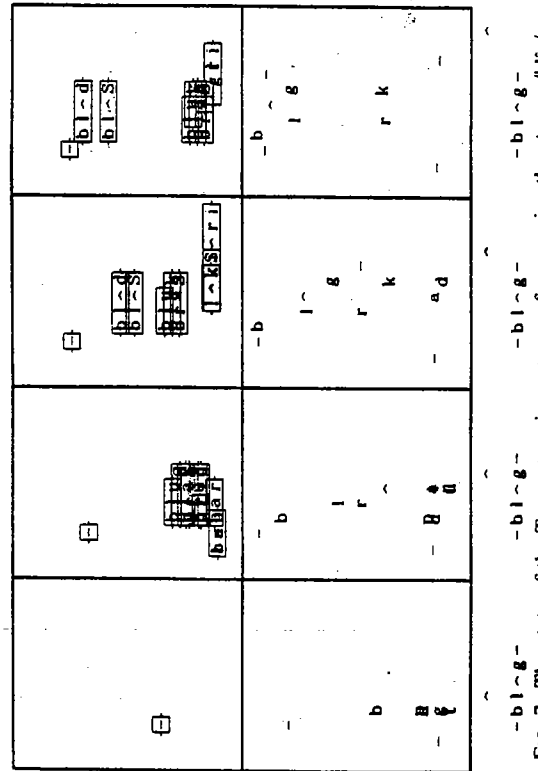


FIG. 7. The state of the Trace at various stages of processing the stream /b'g/.

However, when the ambiguous segment comes late in the word, and the information that precedes the ambiguous segment has already established which of the two alternatives for the ambiguous segment is correct, TRACE shows a lexical effect that develops as the direct perceptual information relevant to the identity of the target segment is being processed. This phenomenon is illustrated in Fig. 8, which shows the state of the Trace at several points in time relative to an ambiguous final segment that could be a /t/ or a /d/, at the end of the context /targ'/. Within the duration of a single phoneme after the center of the ambiguous segment, /t/ already has an advantage over /d/. We therefore predict that Fox's results would come out differently, were he to use word-final, as opposed to word-initial, ambiguous segments. In such a case we would expect the lexical effect to show up well within the 500-ms deadline.

*Dependence of the lexical effect on phonological ambiguity.* One further aspect of the lexical effect that was noted by Ganong (1980) deserves comment. This is the fact that the lexical effect on the identity of a phoneme only occurs with segments which fall in the boundary region between two phonemes. For segments which are unambiguous examples of one category or the other, the effect is not obtained. TRACE is entirely consistent with this aspect of the data. The influence of the lexicon is simply another source of evidence, like that coming from the feature

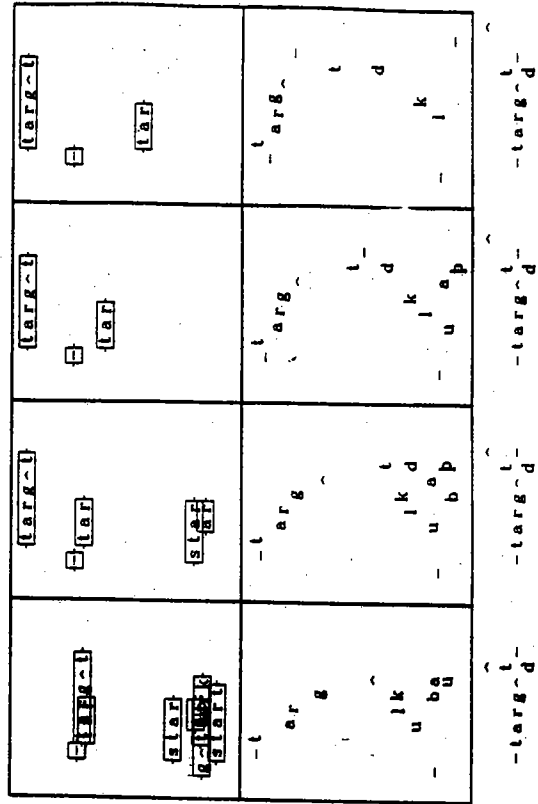


FIG. 8. The state of the Trace at several stages of processing the stream consisting of /targ'/' followed by a segment ambiguous between /t/ and /d/.

level, influencing the activation of one phoneme unit or another. When the bottom-up input is decisive, it can preempt any lexical bias effects. We have verified this in simulations presenting unambiguous tokens of /p/ or /b/, followed either by /l'g/ or /l'S/. In these simulations, the unit for the presented initial segment reaches a very high level of activation, independent of the following context. When the segment comes at the end of the word, the context exerts stronger effects, thus accounting for the fact that speech distortions are easier to detect when they come early in a word than when they come late (Marlsen-Wilson & Welsh, 1978). However, even there, it is possible to override lexically based activations with clear bottom-up signals, although there may be some slowing of the activation process which would probably show up in reaction times.

It should be noted that TRACE's account of lexical effects is quite similar to the account offered by the feature integration theory of Massaro and Oden (1980a). Indeed, Massaro and Oden's model provides quantitative fits to Ganong's findings. We will make some mention of the slight differences in quantitative assumptions between the models below. For now, we note a more crucial difference: TRACE incorporates specific assumptions about the time course of processing which allows it to account for the conditions under which lexical effects will be obtained, as well as for the influence (or a lack thereof) of lexical effects on reaction times, to which we now turn.

*Absence of lexical effect in some reaction-time studies.* Foss and Blank (1980) presented some results which seemed to pose a challenge to interactive models of phoneme identification in speech perception. They gave subjects the task of listening to spoken sentences for occurrences of a particular phoneme in word-initial position. Reaction time to press a response key from the onset of the target phoneme was the dependent variable. In one example, the target was /g/ and the sentence was, *At the end of last year, the government...* The subject's task was simply to press the response key upon hearing the /g/ at the beginning of the word *government*.

The principle finding of Foss and Blank's study was that it made no difference whether the target came at the beginning of a word or a nonword. Later studies by Foss and Gernsbacher (1983) indicate that other experiments which have found lexical or even semantic and syntactic context effects on monitoring latencies are flawed, and that monitoring times for word-initial phonemes are primarily influenced by acoustic factors affecting phoneme detectability, rather than lexical, semantic, or syntactic factors.

The conclusion that phoneme monitoring is unaffected by the lexical status of the target-bearing phoneme string seems at variance with the

spirit of the TRACE model, since in TRACE, the lexical level is always involved in the perceptual process. However, we have already seen that there are conditions under which the lexical level does not get much of a chance to exert an effect. In the previous section we saw that there is no lexical effect on identification of ambiguous word-initial targets when the subject is under time pressure to respond quickly, simply because the subject must respond before information is even available that would allow the model—or any other mechanism—to produce a lexical effect.

In the Foss and Blank situation, there is even less reason to expect a lexical effect, since the target is not an ambiguous segment. We already saw that activation curves rise rapidly for unambiguous segments; in the present case, they can reach near-peak levels well before the acoustic information that indicates whether the target is in a word or nonword has reached the subject's ear.

The results of a simulation run illustrating these points are shown in Fig. 9. For this example, we imagine that the target is /t/. Note how during the initial syllable of both streams, little activation at the word level has been established. Even toward the end of the stream, where the information is just coming in which determines that "trugus" is not a word, there is little difference, because in both cases, there are several active word-level candidates, all supporting the word-initial /t/. It is only after the end of the stream that a real chance for a difference has occurred. Well before this time arrives, the subject will have made a response, since the strength of the /t/ response reaches a level sufficient to guarantee a high accuracy by about Cycle 30, well before the end of the word, as illustrated in Fig. 10.

Even though activations are quite rapid for unambiguous segments, these can still be influenced by lexical effects, provided that the lexical information is available in time. In Fig. 11, we illustrate this point for the phoneme /t/ in the streams /sikr't/ (the word "secret") and /g'd't/ ("guldu't," a nonword). The figure shows the strength of the /t/ response as a function of processing cycles, relative to all other responses based on activations of phoneme units at Cycle 42, the peak of the input specification for the /t/. Clearly, response strength grows faster for the /t/ in /sikr't/ than for the /t/ in /g'd't/; picking an arbitrary threshold of .9 for response initiation, we find that the /t/ in /sikr't/ reaches criterion about 3 cycles or 75 ms sooner than the /t/ in /g'd't/.

*Studies showing lexical effects in reaction times.* Marlsen-Wilson (1980) has reported an experiment that demonstrates the existence of lexical effects in phoneme monitoring for phonemes coming at later points in words. For phonemes coming at the beginning of a word or at the end of the first syllable, he found no facilitation for phonemes in words rel-

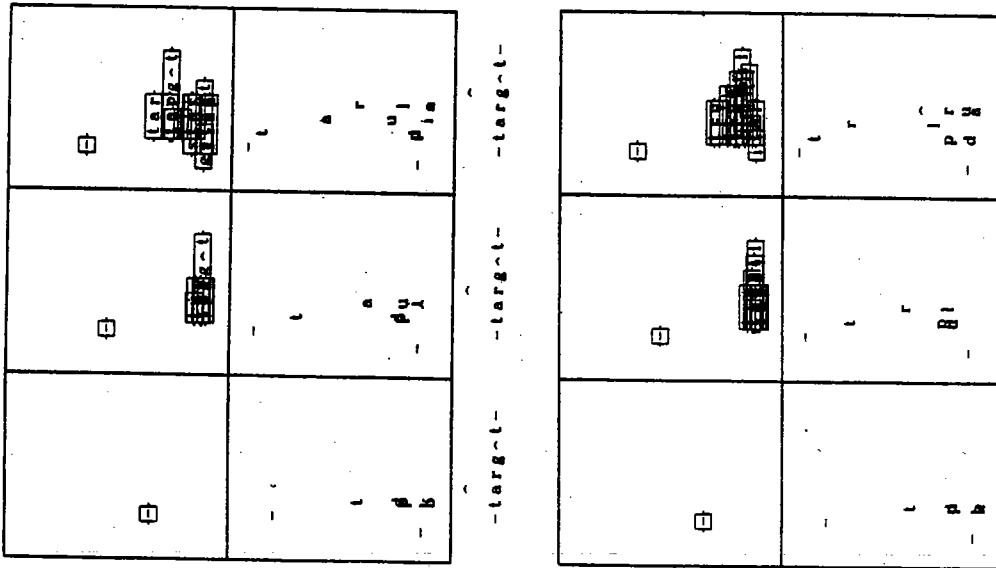


FIG. 9. State of the Trace at three different points during the processing of the word "target" (/targ t/) and the nonword "trugus" (/tr'g's/).

ative to phonemes in nonwords (in fact there was a nonword advantage for these early target conditions). For targets occurring at the end of the second syllable of a two-syllable word (like "secret"—though the stimuli in this particular experiment were Dutch) Marslen-Wilson found an 85-

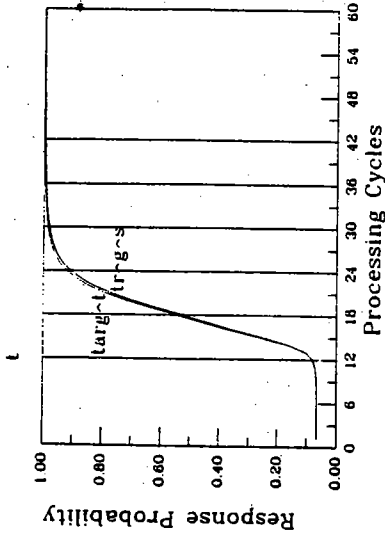


FIG. 10. Time course of growth in the probability of the /t/ response based on activations of phoneme units in Slice 12, during processing of /targ t/ and /tr'g's/. The vertical lines indicate the peaks on the feature patterns corresponding to the successive phonemes of the presented word.

ms advantage compared to corresponding positions in nonwords. This compares quite closely with the value of about 75 ms we obtained for the /sikr't/-g'ld't/ example. At the ends of even longer words, the word advantage increased in size to 185 ms. Marslen-Wilson's result thus confirms that there are indeed lexical effects in phoneme monitoring—even for unambiguous inputs—but underscores the fact that there is no word advantage for phonemes whose processing can be completed long before lexical influences would have a chance to show up.

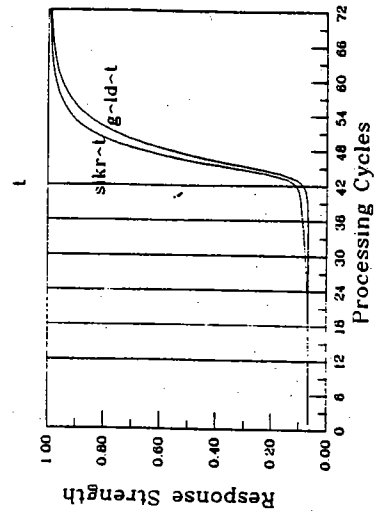


FIG. 11. Probability of the /t/ response as a function of processing cycles, based on activation of phoneme units at Cycle 42, for the stream /sikr't/ ("secret") and /g'ld't/ ("guldu't"). Vertical lines indicate the peaks of the input patterns corresponding to the successive phonemes in either stream.

partial activations of a number of words ("sleep" and "sleet" in the model's lexicon; it would also activate "sleeve," "sleek," and others in a model with a fuller lexicon). None of these word units gets as active as it would if the entire word had been presented. However, all of them (in the simulation, there are only two, but the principle still applies) are partially activated, and all conspire together and contribute to the activation of //l/. This feedback support for the //l/ allows it to dominate the /r/, just as it would if /sli/ were an actual word, as shown in Fig. 12.

The hypothesis that phonotactic rule effects are really based on word activations leads to a prediction: that we should be able to reverse these effects if we present items that are supported strongly by one or more lexical items even if they violate phonotactic rules. A recent experiment by Elman (1983) confirms this prediction. In this experiment, ambiguous phonemes (for example, halfway between /b/ and /d/) were presented in three different types of contexts. In all three types, one of the two (in this case, the /d/) was phonotactically acceptable, while the other (the /b/) was not. However, the contexts differed in their relation to words. In one case, the legal item actually occurred in a word ("bwindle" - "dwindle"). In a second case, neither item made a word, but the illegal item was very close to a word ("bwacelet" - "dwacelet"). In a third case, neither item

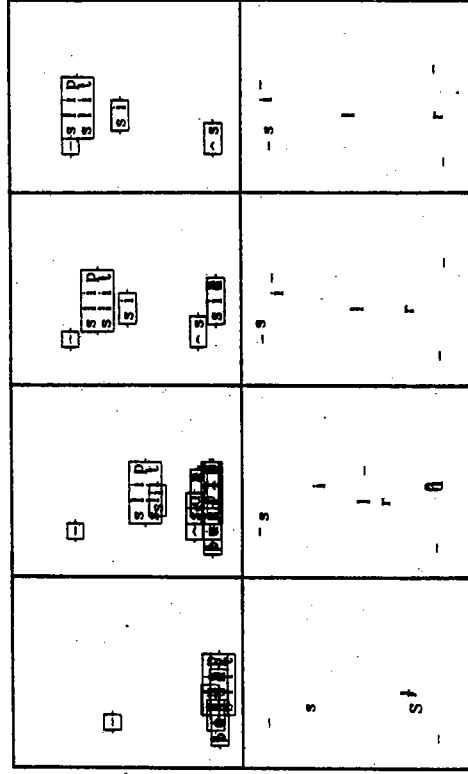


FIG. 12. State of the Trace at several points in processing a segment ambiguous between //l/ and /r/ in the context /s.li/. The units for "sleep" (/sli/) and "sleet" (/sli/) are boxed together since they take on identical activation values.

The TRACE model and Marslen-Wilson's COHORT model (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978) offer fairly similar interpretations of lexical effects in phoneme monitoring. Both models account for the growth in the effect as a function of position in the word. As in COHORT, lexical effects in TRACE depend on the point at which the pattern of activation at the word level begins to specify the identities of the phonemes. In COHORT, there is a discrete moment when this occurs—when the cohort of items consistent with the input is reduced to a single item. In TRACE, things are not quite so discrete. However, it will still generally be the case in TRACE that the size of the lexical effect will vary with the location of the "unique point," the point at which the bottom-up input remains consistent with only a single word. However, since Marslen-Wilson's experiments were performed with Dutch words, we have not been able to simulate his experimental demonstration of this effect in detail.

TRACE and COHORT make similar predictions in some situations, but not in all. In the next section, we consider a phenomenon which TRACE accounts for via the same mechanisms it uses to account for the lexical effects we have been considering. Here, the graded feedback from the word level to the phoneme level allows TRACE to account for an effect that would not be predicted by COHORT, unless additional assumptions were made.

*Are Phonotactic Rule Effects the Result of a Conspiracy?*

Recently, Massaro and Cohen (1983) have reported evidence they take as support for the use of phonotactic rules in phoneme identification. In one experiment, Massaro and Cohen's stimuli consisted of phonological segments ambiguous between /r/ and //l/ in different contexts. In one context (h.li) /r/ is permissible in English, but //l/ is not. In another context (s.li) //l/ is permissible in English but /r/ is not. In a third context (f.li) both are permissible, and in a fourth (v.li) neither is permissible. Massaro and Cohen found a bias to perceive ambiguous segments as /r/ when /r/ was permissible or as //l/ when //l/ was permissible. No bias appeared in either of the other two conditions.

With most of these stimuli, phonotactic acceptability is confounded with the actual lexical status of the item; thus /fri/ and /fri/ ("flee" and "free") are both words, as is /tri/ but not /tli/. In the /s.li/ context, however, neither /sli/ or /sri/ are words, yet Massaro and Cohen found a bias to hear the ambiguous segment as //l/, in accordance with phonotactic rules.

It turns out that TRACE produces the same effect, even though it lacks phonotactic rules. The reason is that the ambiguous stimulus produces

was particularly close to a word ("bwiffle"—"dwiffle"). Results of the experiment are shown in Table 4. The existence of a word identical to one of the two alternatives or differing from one of the alternatives by a single phonetic feature of one phoneme strongly influenced the subject's choices between the two alternatives. Indeed, in the case where the phonotactically irregular alternative ("bwacelet") was one feature away from a particular lexical item ("bracelet"), subjects tended to hear the ambiguous item in accord with the similar lexical item (that is, as a /b/) even though it was phonotactically incorrect.

To determine whether the model would also produce such a reversal of the phonotactic rule effects with the appropriate kinds of stimuli, we ran a simulation using a simulated input ambiguous between /p/ and /t/ in the context /\_Jul/. /p/ is phonotactically acceptable in this context, but /t/ in this context makes an item that is very close to the word "truly." The results of this run, at two different points during processing, are shown in Fig. 13. Early on in processing, there is a slight bias in favor of the /p/ over the /t/, because at first a large number of /p/ words are slightly more activated than any words beginning with /t/. Later, though, the /t/ gets the upper hand as the word "truly" comes to dominate at the word level. Thus, by the end of the word or shortly thereafter, the closest word has begun to play a dominating role, causing the model to prefer the phonotactically inappropriate interpretation of the ambiguous initial segment.

Of course, at the same time the word "truly" tends to support /r/ rather than /l/ for the second segment. Thus, even though this segment is not ambiguous, and the /l/ would suppress the /r/ interpretation in a more neutral context, the /r/ stays quite active.

#### Trading Relations and Categorical Perception

In the simulations considered thus far, phoneme identification is influenced by two different kinds of factors, featural and lexical. When one sort of information is lacking, the other can compensate for it. The image

TABLE 4  
Percentage Choice of Phonotactically Irregular Consonant

Stimulus type	Example	Percentage of identifications as "illegal" phoneme <sup>a</sup>
Legal word/illegal nonword	dwindle/bwindle	37
Legal nonword/illegal nonword	dwiffle/bwiffle	46
Legal nonword/illegal nearword	dwacelet/bwacelet	55

<sup>a</sup>  $F(2,34) = 26.414, p < .001$ .

that emerges from these kinds of findings is of a system that exhibits great flexibility by being able to base identification decisions on different sources of information. It is, of course, well established that within the featural domain each phoneme is generally signaled by a number of different cues, and that human subjects can trade these cues off against each other. The TRACE model exhibits this same flexibility, as we detail shortly.

But there is something of a paradox. While the perceptual mechanisms exhibit great flexibility in the cues that they rely on for phoneme identification, they also appear to be quite "categorical" in nature. That is, they produce much sharper boundaries between phonetic categories than we might expect based on their sensitivity to multiple cues, and they appear to treat acoustically distinct feature patterns as perceptually equivalent, as long as they are identified as instances of the same phoneme.

In this section, we illustrate that in TRACE, just as in human speech perception, flexibility in feature interpretation—specifically, the ability to trade one feature of a phoneme off against another—coexists with a strong tendency toward categorical perception.

For these simulations, the model was stripped down to the essential minimum necessary, so that the basic mechanisms producing cue trade-

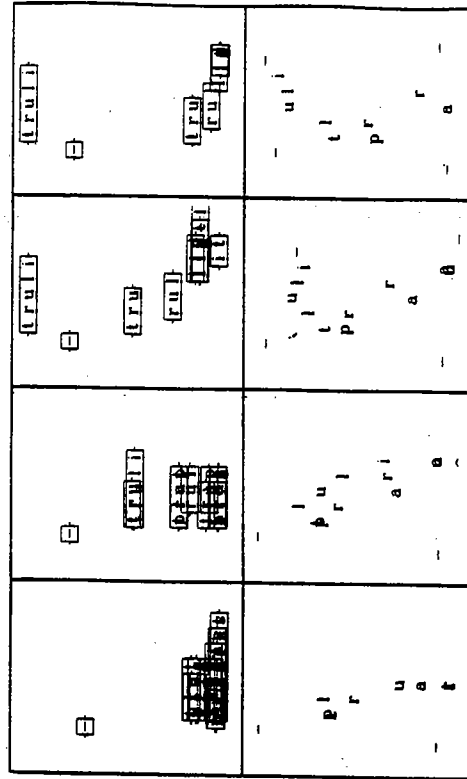


Fig. 13. State of the Trace at several points in processing an ambiguous /p/-/t/ segment followed by /luli/.

several categorical perception studies of VOT continua (using /g-/k/, /d-/t/, or /b-/p/ stimuli) have covaried both VOT and FIOF, if only because FIOF tends to covary with VOT when realistic stimuli are used (e.g., Pisoni & Lazarus, 1974; Samuel, 1977). Though the simulations use a /g-/k/ continuum, we consider several categorical perception experiments using /d-/t/ and /b-/p/ continua, since the same dimensions can differentiate the two members of both of these other pairs. We also consider data obtained in experiments on other continua, using other cues. We could easily have repeated the simulations with other sets of continua; however, the general qualitative form of the results would be the same. What would vary from case to case would be the magnitude of the effect of a step along a given dimension.

The pattern of excitatory input to the VOT and FIOF detectors produced by the canonical mock speech /g/ and /k/ used in the simulations are illustrated in Fig. 15.

*Trading relations.* TRACE quite naturally tends to produce trading relations between features, since it relies on the weighted sum of the excitatory inputs to determine how strongly the input will activate a particular phoneme unit. All else being equal, the phoneme unit receiving the largest sum bottom-up excitation will be more strongly activated than any other, and will therefore be the most likely response when a choice must be made between one phoneme and another. Since the net bottom-up input is just the sum of all of the inputs, no one input is necessarily decisive in this regard.

Generally, experiments demonstrating trading relations between two or more cues manipulate each of the cues over a number of values ranging between a value more typical of one of two phonemes and a value more typical of the other. Summerfield and Haggard did this for VOT and FIOF, and found the typical result, namely that the value of one cue that gives rise to 50% choices of /k/ was affected by the value of the other cue: the higher the value of FIOF, the shorter the value of VOT needed for 50% choices of /k/. Unfortunately, they did not present full curves relating phoneme identification to the values used on each of the two dimensions. In lieu of this, we present curves in Fig. 16 from a classic trading relations experiment, by Denes (1955). Similar patterns of results have been reported in other studies, using other cues (e.g., Massaro, 1981, Figs. 4 and 5), though the transitions are often somewhat steeper (see below for a discussion of the issue of steepness). We have chosen to present the shallower curves reported by Denes because in them we see clearly that there are cases in which a cue that favors one of the two phonemes to a moderate degree will give rise to the perception of the other phoneme when paired up with a strong cue that favors the other

offs and categorical perception could be brought to the fore. The word level was eliminated altogether, and at the phoneme level there were only three phonemes, /a/, /g/, and /k/, plus silence (-). From these four items, inputs and percepts of the form /-ga-/ and /-ka-/ could be constructed. The following additional constraints were imposed on the feature specifications of each of the phonemes: (1) the /a/ and /-/ had no overlap with either /g/ or /k/, so that neither /a/ nor /-/ would bias the activations of the /g/ and /k/ phoneme units where they overlapped with the consonant; (2) /g/ and /k/ were identical on five of the seven dimensions, and differed only on the remaining two dimensions.

The two dimensions which differentiated /g/ and /k/ were voice onset time (VOT) and the onset frequency of the first formant (FIOF). These dimensions replaced the voicing and burst amplitude dimensions used in all of the other simulations. Figure 14 illustrates how FIOF tends to increase as voice onset time is delayed.

Summerfield and Haggard (1977) have shown that subjects are sensitive both to VOT and to FIOF and that it is possible to trade one of these cues off against the other. Thus, the boundary between /ga/ and /ka/ shifts to longer VOTs when F1 starts off lower rather than higher.

Categorical perception and trading relations among cues have been studied on a variety of different continua by a variety of different investigators. We have chosen to focus on the VOT and FIOF features, as exemplified by the /ga-/ka/ continuum, because there is data on trade-offs between these cues (Summerfield & Haggard, 1977), and because

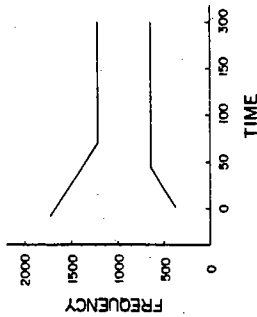


FIG. 14. Schematic diagram of a syllable that will be heard as /ga/ or /ka/, depending on the point in the syllable at which voicing begins. Prior to the onset of voicing, F2 (top curve) is energized by aperiodic noise sources, and F1 is "cut back" (the noise source has little or no energy in this range). Because of the fact that F1 rises over time after syllable onset (as the vocal tract moves from a shape consistent with the onset of voicing to a shape consistent with the vowel), its frequency at the onset of voicing is higher for later values of VOT. Parameters used in constructing this schematic syllable are derived from Kewley-Port's (1982) analysis of the parameters of formants in natural speech, and are similar to those used in many perceptual experiments.

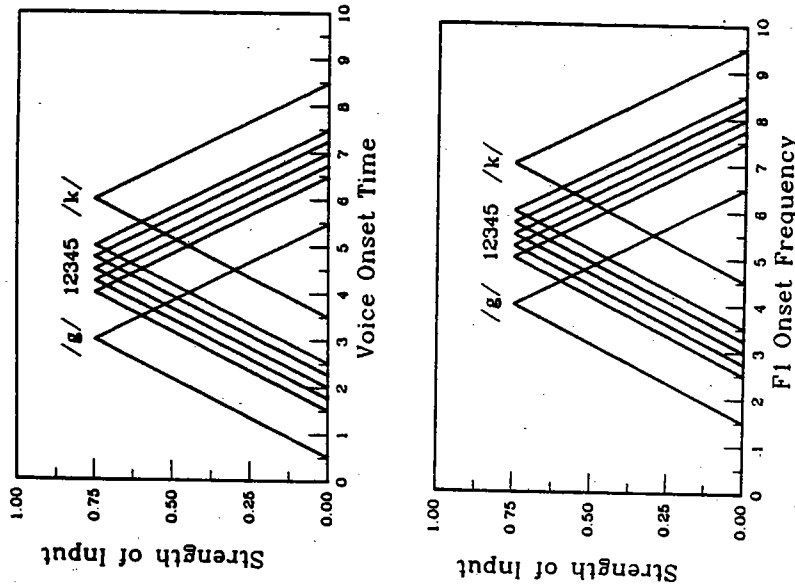


FIG. 15. Canonical feature-level input for /g/ and /k/, on the two dimensions that distinguish them, and the patterns used for the five intermediate values used in the trading relations simulation. Along the abscissa of each dimension the nine units for the nine different value ranges of the dimension are arrayed. The curves labeled /g/ and /k/ indicate the relative strength of the excitatory input to each of these units, produced by the indicated phoneme. The canonical curves also indicate the strengths of the feature-to-phoneme connections for /g/ and /k/ on these dimensions. That is, the canonical input pattern for each phoneme exactly matches the strengths of the corresponding feature-phoneme connections. Numbered curves on each dimension show the feature patterns used in the trading relations simulation.

phoneme. An additional finding is the bowing of the curves; they tend to be approximately linear through the middle of their range, but to level off at both ends, where the values on both dimensions agree in pointing to one alternative or the other.

To see if TRACE would simulate the basic trade-off effect obtained by Summerfield and Haggard, and to see if it would produce the same shape

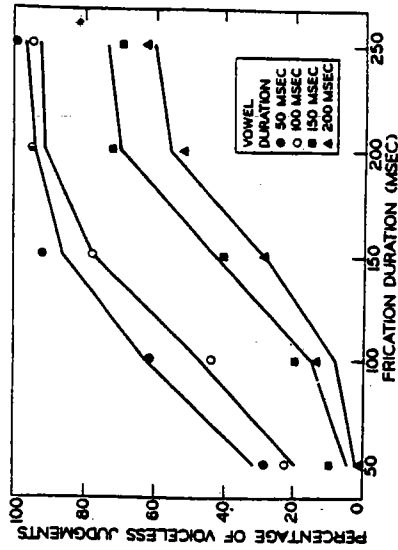


FIG. 16. Results of an experiment demonstrating the trade-off between two cues to the identity of /s/ and /z/. Data from Denes, 1955, fitted by the model of Massaro and Cohen, 1977. ●, 50 ms; ○, 100 ms; ■, 150 ms; ▲, 200 ms. Reprinted with permission from Massaro and Cohen (1977).

trade-off curves as have been generally reported, we generated a set of 25 intermediate phonetic segments made up by pairing each of five different intermediate patterns on the VOT dimension with each of five different intermediate patterns on the FIOF dimension. The different feature patterns used on each dimension are shown in Fig. 15, along with the canonical feature patterns for /g/ and /k/ on each of the two dimensions. On the remaining five dimensions, the intermediate segments all had the common canonical feature values for /g/ and /k/.

The model was tested with each of the 25 stimuli, preceded by silence (/—/) and followed by /a—/. In this and all subsequent simulations we report in this paper, the peak of the initial silence phoneme occurred at Time Slice 6 in the input, and the peaks of successive phoneme segments occurred at six slice intervals. Thus, for these stimuli, the peak on the intermediate phonetic segment occurred at Slice 12, the peak of the following vowel occurred at Slice 18, and the peak of the final silence occurred at Slice 24. For each input presented, the interactive activation process was allowed to continue through a total of 60 time slices, well past the end of the input. The state of the Trace at various points in processing, for the most /g/-like of the 25 stimuli, is shown in Fig. 17. At the end of the 60th time slice, we recorded the activation of the units for /g/ and /k/ in Time Slice 12 and the probability of choosing /g/ based on these activations. (It makes no difference to the qualitative appearance of the results if a different decision time is used; earlier decision times are associated with smaller differences in relative activation between the /g/ and /k/ phoneme units, and later ones with larger differences, but the general pattern is the same.)

/k/-like values on both dimensions. In terms of Summerfield and Haggard's measure, the value of VOT needed to achieve 50% probability of reporting /k/, we can see that the VOT needed increases as the FIOF decreases, just as these investigators found.

Cue trade-offs in phoneme identification are accounted for in detail by the feature integration model of Oden and Massaro (1978; Massaro, 1981; Massaro and Oden, 1980a, 1980b). While we have shown how TRACE can account for the basic trade-off effect and the general form of the trade-off curves, we have not yet attempted the kinds of detailed fits that Massaro, Oden, and collaborators have reported in a number of studies. However, the models are quite similar, so it seems rather unlikely that cue trade-off data would be able to discriminate between them. And both make special assumptions about lack of invariance of cues to phoneme identity across contexts.

One apparent dissimilarity between the models deserves comment. Whereas cue strengths are combined multiplicatively in the determination of response strengths in the feature integration model, they are combined additively in the bottom-up inputs to the units in TRACE. However, in TRACE, two further computational steps take place before these inputs result in response strengths. First, the interactive-activation process enhances differences between competing units. Second, the resulting unit activations are subjected to an exponential transformation. Just this second step by itself would transform influences that have additive effects on unit activations into influences that have multiplicative effects on response strength. Thus, the models would be mathematically equivalent if the interactive activation process were simply replaced by a linear, additive combination of inputs to the units. In quantitative formulations of the interactive activation process closely related to the ones we use (Grossberg, 1978), what the interactive activation process does is simply rescale the unit activations, preserving the ratios of their bottom-up activation but keeping them bounded. Though our version of these equations does not do this exactly, the ways in which it deviates from this would be difficult to use as the basis for an empirical distinction between the TRACE approach and the feature integration model. Thus, up to a point, we can see TRACE as (approximately) implementing the computations specified in Oden and Massaro's model. The models differ, though, in that TRACE is dynamic and in that it incorporates feedback to the phoneme level. This allows TRACE to account for categorical perception in a different way.

*Categorical perception.* In spite of the fact that TRACE is quite flexible in the way it combines information from different features to determine the identity of a phoneme, the model is quite categorical in its overt responses. This is illustrated in two ways: first, the model shows a much sharper transition in its choices of responses as we move from /g/ to /k/

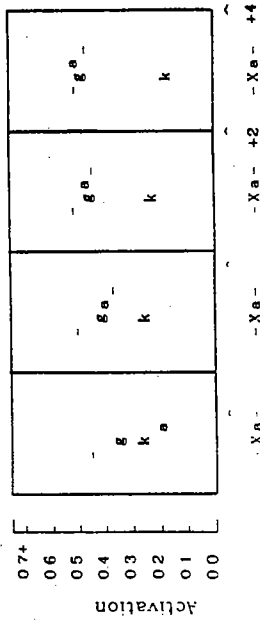


FIG. 17. The state of the Trace at various points during and after the presentation of a syllable consisting of the most /g/-like of the 25 intermediate segments used in the trading relations experiment, represented by /X/, preceded by silence and followed by /a/. then another silence.

Response probabilities were computed using the formulas given earlier for converting activations to response strengths and strengths into probabilities. The resulting response probabilities, for each of the 25 conditions of the experiment, are shown in Fig. 18. The pattern of results is quite similar to that obtained in Denes (1955) experiment on the /s/-/z/ continuum. The contribution of each cue is approximately linear and additive in the middle of the range, but the curves flatten out at the extremes, as in the Denes (1955) experiment. More importantly, the model's behavior exhibits the ability to trade one cue off against another. For example, there are three different combinations of feature values which lead to a probability between .82 and .85 of choosing /k/: (1) the neutral value of the VOT dimension coupled with the most /k/-like value on the FIOF dimension; (2) the neutral value on the FIOF dimension coupled with the most /k/-like value of the VOT dimension; and (3) the somewhat

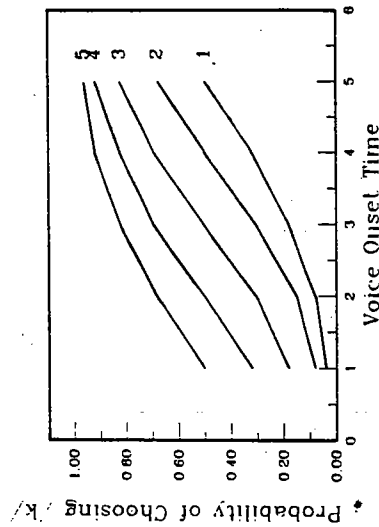


FIG. 18. Simulated probability of choosing /k/ at Time Slice 60, for each of the 25 stimuli used in the trading relations simulation experiment. Numbers next to each curve refer to the intermediate pattern on the FIOF continuum used in the five stimuli contributing to each curve. Higher numbers correspond to higher values of FIOF.

along the VOT and FIOF dimensions than we would expect from the slight changes in the relative excitation of the /g/ and /k/ units. Second, the model tends to obliterate differences between different inputs which it identifies as the same phoneme, while sharpening differences between inputs assigned to different categories. We will consider each of these two points in turn, after we describe the stimuli used in the simulations.

Eleven different consonant feature patterns were used, embedded in the same simulated /-a-/ context as in the trading relations simulation. The stimuli varied from very low values of both VOT and FIOF, more extreme than the canonical /g/, through very high values on both dimensions, more extreme than the canonical /k/. All the stimuli were spaced equal distances apart on the VOT and FIOF dimensions. The locations of the peak activation values on each of these two continua are shown in Fig. 19.

Figure 20 indicates the relative initial bottom-up activation of the /g/ and /k/ phoneme units for each of the 11 stimuli used in the simulation. The first thing to note is that the relative bottom-up excitation of the two phoneme units differ only slightly. For example, the canonical feature pattern for /g/ sends 75% as much excitation to /g/ as it sends to /k/. The feature pattern two steps toward /g/ from /k/ (Stimulus 5), sends 88% as much activation to /g/ as to /k/.

The figure also indicates, in the second panel, the resulting activations

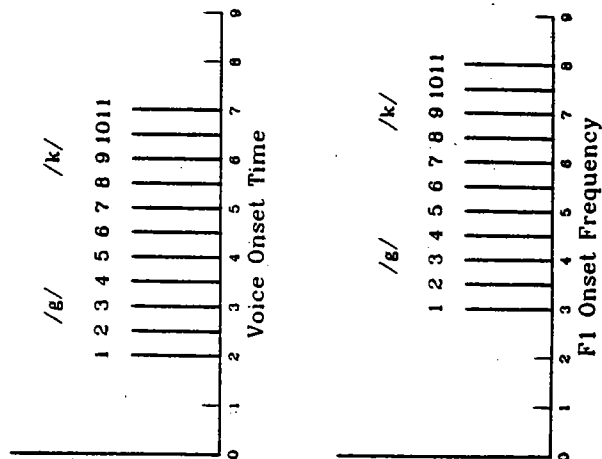


Fig. 19. Locations of peak activations along the VOT and FIOF dimensions, for each of the 11 stimuli used in the categorical perception simulation.

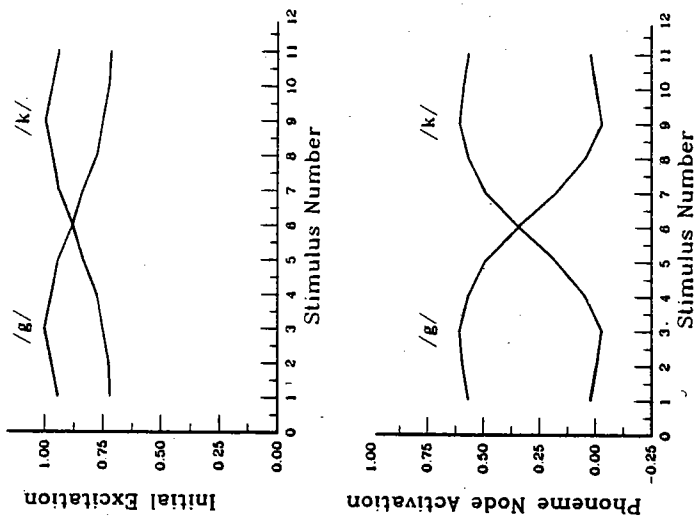


Fig. 20. Effects of competition on phoneme activations. The first panel shows relative amounts of bottom-up excitatory input to /g/ and /k/ produced by each of the 11 stimuli used in the categorical perception simulation. The second panel shows the activations of units for /g/ and /k/ at Time Cycle 60. Stimuli 3 and 9 correspond to the canonical /g/ and /k/, respectively.

of the units for /g/ and /k/ at the end of 60 cycles of processing. The slight differences in net input have been greatly amplified, and the activation curves exhibit a much steeper transition than the relative bottom-up excitation curves.

There are two reasons why the activation curves are so much sharper than the initial bottom-up excitation functions. The primary reason is *competitive inhibition*. The effect of the competitive inhibition at the phoneme level is to greatly magnify the slight difference in the excitatory inputs to the two phonemes. It is easy to see why this happens. Once one phoneme is slightly more strongly activated than the other, it exerts a stronger inhibitory influence on the other than the other can exert on it. The net result is that "the rich get richer." This general property of competitive inhibition mechanisms was discussed by McClelland and Rumelhart (1981), following earlier observations by Grossberg (see Grossberg, 1978, for a discussion) and Levin (1976); it is also well known as one possible basis of edge enhancement effects in low levels of visual

