

## Cognitive Penetration of the Mechanisms of Perception: Compensation for Coarticulation of Lexically Restored Phonemes

JEFFREY L. ELMAN

*University of California, San Diego*

AND

JAMES L. MCCLELLAND

*Carnegie Mellon University*

An important question in language processing is whether higher-level processes are able to interact directly with lower-level processes, as assumed by interactive models such as the TRACE model of speech perception. This issue is addressed in the present study by examining whether putative *interlevel* phenomena can trigger the operation of *intralevel* processes at lower levels. The intralevel process involved the perceptual compensation for the coarticulatory influences of one speech sound on another. TRACE predicts that this compensation can be triggered by illusory phonemes which are perceived as a result of top-down, lexical influences. In Experiment 1, we confirm this prediction. Experiments 2 to 4 replicate this finding and fail to support several potential alternative explanations of the results of Experiment 1. The basic finding that intralevel phenomena can be triggered by interlevel processes argues against the view that aspects of speech perception are encapsulated in a module impervious to influences from higher levels. Instead, it supports a central premise of interactive models, in which basic aspects of perceptual processing are subject to influences from higher levels. © 1988 Academic Press, Inc.

How far down into the mechanisms of perception do higher-level contextual influences reach? There has been considerable debate on this issue and the more general question of the degree to which cognitive processes interact with one another. Some theorists (e.g., Fodor, 1983) have proposed that processing is essentially modular in nature, and that the notable feature of cognition is the autonomy of mental faculties. This is the *autonomous* or *modular* view. Others have argued that the flow of information is rather freer, and that top-down as well as bottom-up interactions are possible. This is the *interactive* view. The issue is a difficult one to decide, because there is

Requests for reprints should be addressed to Dr. Elman at the Department of Linguistics, University of California, San Diego, La Jolla, CA 92093.

considerable overlap in the predictions made by the two accounts. Many of the phenomena which have been cited as evidence for either the autonomous or interactive theories are in fact compatible with both.

In this paper we consider the question as it arises in the perception of speech. We describe a technique which we believe provides a rigorous test of the existence of true top-down interactions in processing. The technique is applied here to the domain of speech perception, but we believe it can also be used in other areas where these issues arise. Using this technique, we find that higher-level contextual factors can trigger compensatory processes that are basic to speech perception. These findings demonstrate that a more direct effect of

higher-level knowledge on perception is possible than has been thought.

#### EVIDENCE FOR INTERACTIONS

The degree to which cognitive processes are modular has been an issue of great interest and controversy. At first glance, the case for interactions, and, in particular, top-down effects, might seem to be rather strong. Marslen-Wilson, Tyler, and their colleagues (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978) have shown that language comprehension appears to involve the simultaneous processing of information at the semantic, syntactic, and phonological levels (at least). Furthermore, this processing seems to be not only parallel but interactive; that is, processing at each level appears to influence, and be influenced by, processing at other levels (cf. Rumelhart, 1977). The result is that many tasks, such as error detection and shadowing, are accomplished more quickly when information from these various levels is simultaneously available (Cole, 1973; Cole & Jakimik, 1978, 1980; Cole, Jakimik, & Cooper, 1978). Warren and Sherman (1974) have demonstrated that lexical information appears to play an active role in the perception of phonemes; they obtain a "phoneme restoration" effect in which obliterated phonemes are perceived as being present, when the context is a lexical item. Ganong (1980) has shown that the perceptual boundary between phonemes can be shifted as a function of lexical context; thus, a sound acoustically midway between [g] and [k] is perceived as a [k] in the context "-iss," but as a [g] in the context "-ift." Other evidence supporting the interactionist position on lexical processing has been presented by Morton (1969), Grosjean (1980), and others (see also Samuel, 1986, and McClelland & Elman, 1986, for extensive considerations of additional literature).

On the other hand, it has been claimed that processing is actually autonomous and that bottom-up processes make their

output available to higher-level processes, but that the latter do not affect the operation of the former (e.g., Forster, 1979; Tanenhaus, Carlson, & Seidenberg, 1984). In this view, lower-level processes would extract cues to the identity of speech sounds from the acoustic signal. This part of the mechanism is sensitive to local acoustic contextual influences but not to higher-level contextual factors, such as whether a sequence of phonemes forms a word or is consistent with the more global context. A second stage would take the output of the first stage and combine this output with higher-level factors, such as knowledge of what phonemes make a word, to make decisions about the identity of phonemes.

There are several ways in which the autonomous model can explain the effects that would seem to argue for freer interactions. Norris (1982) has proposed that the output of the lower-level processes may be quite detailed. If the early stages of processing are unable to resolve ambiguous input, they simply need to provide a rich enough output that later stages can make the correct interpretation (based on information which is available to them, but not to the lower stages). Thus, for example, the lexical bias on phoneme identification discovered by Ganong (1980) would be explained as the effect of a postperceptual response strategy which results in responses that are more consistent with lexical items.

Another account of how some apparent top-down interactions can arise in a strictly bottom-up model is suggested by Fodor (1983) and by Forster (1981). It is plausible that frequent sequences of events would result in stronger intramodule associations for these sequences. Thus, frequent sequences of words or frequent sequences of phonemes might facilitate filling in of degraded or missing input; but the filling in would arise from intramodule dynamics and not from higher-level information.

It is thus difficult to know whether, in a given instance, the effects of context are

due to intralevel dynamics or to higher-level postperceptual adjustments or to true top-down interactions. What would such a test involve?

In order to answer that question, let us first describe a model of speech perception we have developed, called the TRACE model (Elman & McClelland, 1986; McClelland & Elman, 1986). This model holds that there are interactions between the syntactic/semantic level of processing and the lexicon and between the lexicon and phonetic analysis, with information flowing in both directions between adjacent levels. TRACE accounts for a large body of empirical data. For current purposes, however, we want to describe how the model suggests a method for addressing the issue of interactions between levels. In so doing, TRACE not only provides an account for existing data, but it also serves as the driving force for new empirical findings.

#### THE TRACE MODEL OF SPEECH PERCEPTION

The TRACE model is based on an interactive-activation model of processing (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1981). Processing is carried out in a network consisting a large number of interconnected elements called units. Different classes of units represent acoustic/phonetic features, phonemes, and words. At the feature level, there is a unit for each feature at each of a large number of time slices relative to the onset of an utterance. At the phoneme level, there is a unit for each phoneme in each time slice. At the word level, there is a unit for each word, starting in each time slice.

Each unit has an activation value which is taken to be translatable into an estimate of the strength of the hypothesis that the concept the unit represents is present in the signal in the time slice or slices that the unit covers. Thus, the pattern of activation over the units in the network is the system's representation of the content of the utterance it is currently processing. This representa-

tion evolves through time, as activations change in the course of processing.

Activations of units are continuously updated in TRACE, based on activations of units that project to them. When a unit's activation exceeds a threshold value, it excites or inhibits other units in a way which reflects the mutual consistency or inconsistency of the concepts represented by each unit. Influences are bidirectional; feature units excite units for the phonemes that contain them, and phoneme units excite units for the features they contain. Similarly, phoneme units excite units for the words that contain them, and word units excite units for the phonemes they contain. There are also inhibitory connections between mutually incompatible units on the same level. Thus, units for different phonemes in the same time slice are mutually inhibitory, as are units for different words that span overlapping ranges of time slices (here the extent of inhibition is proportional to the degree of overlap).

The activations of units are determined by a simple two-part activation function. The first part consists of adding together all of the inputs to the unit, to obtain what is called its *net input*. This net input is then used to update the activation of the unit. If the net input is positive (excitatory), it tends to increase the activation of the unit; if it is inhibitory, it tends to decrease the unit's activation. The activation function is approximately linear near the middle of its range, but is nonlinear at the extremes, so that activations of units receiving strongly excitatory input level off below the maximum activation of 1.0, and activations of units receiving strongly inhibitory input level off above the minimum activation, which is set at a value slightly below the unit's threshold.

#### *Context Effects in TRACE*

Let us consider for a moment how the TRACE model accounts for context effects on phoneme identification, since the details of this process will be relevant when we

consider how one might test the assumption that influences really do feed back down from higher levels within the speech-processing system. For concreteness, consider the identification of an ambiguous segment, providing equal bottom up input to /b/ and a /p/, followed by /l^g/, so that the whole input is between *blug* and *plug*.

Figure 1 illustrates simulated activations of units at the phoneme and word levels at several points in this process. The input is assumed to unfold in time, with features of the ambiguous segment arriving first. This input activates feature level units, which in turn provide bottom-up input to phoneme units in the corresponding time slices. The greatest amount of excitation is received by units for /b/ and /p/, so the activations of both units begin to build up. By the time the /l/ in the input is coming in (first panel of the figure) these two phonemes have reached an activation of about 0.2, and are beginning to excite units at the word level, though none have exceeded threshold (activation of 0) at this point. As the speech

input continues to unfold, detectors for other phonemes become activated. As phoneme level activations build up, the active phonemes begin to excite units at the word level. At first, all words beginning with /b/ and /p/ receive equal excitation; as the /l/ in the input becomes active words beginning in /bl/ and /pl/ begin to dominate other words, then those beginning in /bl^/ and /pl^/ (including "blood," "blush," "plug," and "plum") begin to win out, as is illustrated in the second panel of the figure. When the final /g/ becomes active at the phoneme level (third panel), /pl^g/ wins out over the others because it receives the most bottom-up support. Up to this point, both the /b/ and the /p/ have been receiving top-down activation from the active units that contain them at the word level, but as /pl^g/ gradually establishes its dominance at the word level, it supports the /p/ at the phoneme level and inhibits all other word units, thereby removing the top-down support for the /b/. The result is eventual dominance of /p/ over /b/ (final panel of the

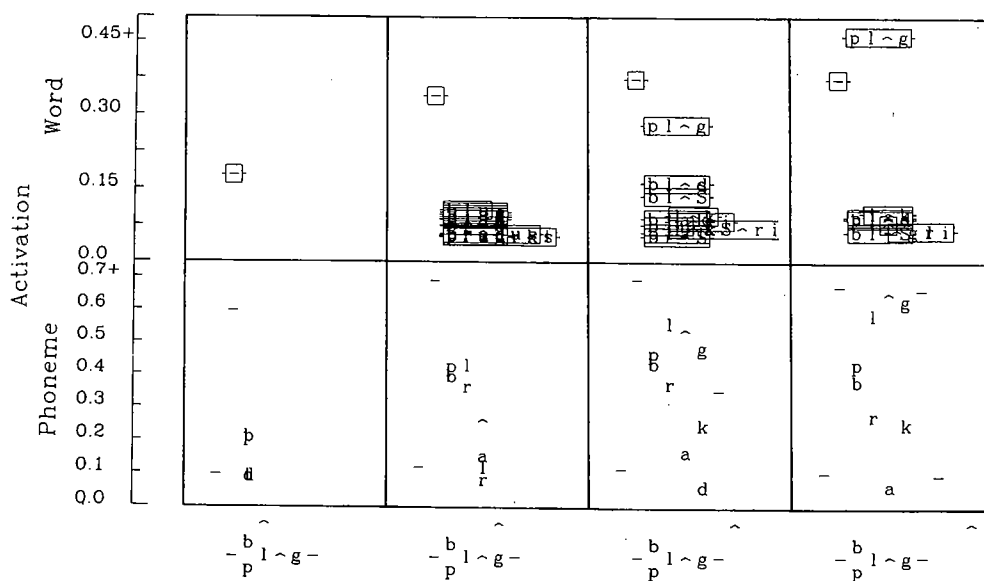


FIG. 1. The time course of the build up of strength of the /p/ response based on activations of phoneme units in Slice 12, in processing an ambiguous /b/-/p/ segment followed by /l^g/ (as in "plug") and the same segment followed /l^s/ (as in "blush"). The ambiguous segment is indicated by the "?". Also shown is the build up of response strength for an unambiguous /p/ segment in /pl^g/. The vertical line topped with "?" indicates the point in processing corresponding to the center of the initial segment in the input stream. Successive vertical lines indicate centers of successive phonemes. (From The TRACE model of Speech Perception, by J. L. McClelland & J. L. Elman, 1986, *Cognitive Psychology*, 18, 1-86. Reprinted by permission.)

figure). The effect is not terribly dramatic in this case, in large part because the context does not come in until after the ambiguous segment; and the effect does not become apparent until after the final phoneme has been heard.

Like other models, TRACE does require a readout process to translate activations of units into overt responses. TRACE's assumptions about the readout process are shared with the interactive activation model, and derive ultimately from the choice model of Luce (1959; see McClelland & Elman, 1986, for more details). For present purposes, it is sufficient to note that the readout assumptions of TRACE and the interactive-activation model allow the models to approximate the quantitative predictions of models in which information from lexical and featural inputs is integrated in a postperceptual decision mechanism of the kind described by Massaro (1979, 1988).<sup>1</sup>

We have examined this example in some detail, not only to give the reader a sense of the course of events in TRACE, but also to bring out two features of its operation that are relevant to the issue of testing the claim, inherent in the model, that there is feedback from the word level to the phoneme level. These features are as follows: (1) Top-down input influences processing by combining additively with bottom-up input in determining the total, or what is often called the *net* input to the unit. Thus activation from the word level is treated as just another source of excitatory input, indistinguishable from bottom-up input in kind. (2) The top-down information process influences the activation of a phoneme unit takes time. This is particularly true when the context that determines the identity of the initial phoneme comes

after the phoneme, but even when the context comes before, it can take a while for the effects of this context to percolate up to the word level and back down again (see McClelland, 1987).

With these two points in mind, we can evaluate two different sorts of evidence that have been offered for assessing whether an effect is truly top-down or not. One of these is based on signal detection theory (SDT; Green and Swets, 1966). Within the context of signal detection theory, one may ask, does context affect sensitivity ( $d'$ ) or does it simply bias responses in one direction or another, without altering the inherent perceptibility of speech sounds? At first glance, it would seem that if context only influences bias, that is, the  $\beta$  parameter of SDT, then context is simply operating at a response selection or decision stage. It is true that  $\beta$  can be affected by influences operating on the process of deciding what overt response to make based on the results of perceptual processes. However, it should be noted that in TRACE, the top-down effects we have been examining show up primarily in the bias measure, rather than in  $d'$ , even though they operate within the processing system itself, before the response stage. The reason is that an active word unit simply adds its contribution to the net input of a phoneme unit that it supports. The effect is a true top-down effect, in that context is influencing the very same detectors that are influenced by bottom-up input, in a way that is indistinguishable from the way in which perceptual input influences these detectors. Yet the result does not ordinarily involve any change in sensitivity, but simply pushes the subject along the likelihood dimension in the direction of choosing the contextually appropriate response.

Based on the foregoing, it should be apparent that it is completely consistent with the assumptions of the TRACE model that contextual cues combine with bottom-up cues to phoneme identity in just the same

<sup>1</sup> The correspondence to Massaro's model is not exact. It can be quite close in idealized cases—indeed the mathematical formulations are closely related—but interactions among units between and within levels tend to distort the ideal somewhat.

way that bottom-up cues combine with each other (Massaro, 1979, 1988). Essentially, each cue can be seen as contributing to the total number of votes for one candidate response or another. In some models, such as Massaro's, these votes are combined in a decision process that reads out the results of processing. In TRACE, they are combined within the processing mechanism itself. Indeed, it is possible to view TRACE and related interactive models as incorporating into the internal workings of the perceptual mechanism the fundamental decisional processes described by Massaro and others.<sup>2</sup>

Thus far we have seen that TRACE is consistent with evidence that lexical context often operates like other cues to phoneme identity, producing an effect which can show up as a bias effect in a signal detection analysis. However, this leaves us no better off than we were before, since it indicates that one cannot use the absence of a  $d'$  effect to argue against an interactive position.

Another line of argument that has been offered in support of a postperceptual, response-stage view of context effects in phoneme perception is the fact that they often take time to build up. Fox (1982) demonstrated that subjects who were forced to respond within 500 ms of the onset of an ambiguous phoneme occurring

at the beginning of a word, as in our /ʔl^g/ example above, did not show a context effect; that is, they did not show a bias in favor of the phoneme that makes a word in the context. A context effect occurred only when subjects were not forced to respond under time pressure. On the basis of these findings, Fox argued that the effect was a late, postperceptual one. We have already seen, however, that in TRACE, the effect can be slow to develop, without being any different in kind from the initial bottom-up activation process. Of course, this view predicts that context effects will operate more quickly for word-final than word-initial ambiguities, and though Fox's experiment has not been repeated exactly for this case, this prediction has been confirmed in a slightly different paradigm (Marslen-Wilson, 1980).

We see, then, that previous evidence taken in support of models in which the effects of context operate on decision mechanisms does not in fact constitute evidence that is inconsistent with the TRACE model. In fact, these previous experiments underscore what we take to be a positive feature of TRACE and other interactive models; these models embed the information integration processes traditionally taken as the hallmark of postperceptual decision mechanisms into the perceptual machinery itself. However, we are still left without a definitive test that distinguishes the interactive models from accounts in which context effects arise in a postperceptual decision mechanism operating on the outputs of a strictly bottom-up perceptual processing system.

However, there is one more aspect of TRACE which we have not yet mentioned that provides a basis for finding just such a test. This is the mechanism that TRACE uses to deal with coarticulatory influences in speech. We first describe the coarticulatory influences, and the mechanism TRACE provides to accommodate them. Then we consider how these influences can be used to test TRACE's assumption that

<sup>2</sup> For completeness we note that there are circumstances in which context does produce sensitivity effects; the word superiority effect (Reicher, 1969) is a case in point, and Samuel (1981, 1986) has explored cases of sensitivity effects in speech perception. A discussion of the conditions under which these effects occur, and how they may be accounted for in interactive-activation models, is beyond the scope of the present paper (but see McClelland & Rumelhart, 1981, for a full account for the Reicher effect in the interactive-activation model of visual word processing). Suffice it to say that sensitivity effects cannot be taken as unambiguous evidence of top-down effects. For examples of models that account for sensitivity effects of context on letter identification in the Reicher task without invoking feedback from higher levels to lower levels, see Johnston and McClelland (1980), and Thompson and Massaro (1973).

higher levels to indeed feed activation back to lower levels.

### Compensation for Coarticulation

Coarticulatory influences in speech result from vocal tract dynamics and are a major source of variability in the speech signal. It is well known that listeners compensate perceptually for the coarticulatory influences of one phoneme on the production—and hence the acoustic realization—of its neighbors (Repp & Liberman, 1984). Indeed, adjustment to such coarticulatory influences is thought to be among the most essential tasks of the early stages of processing in speech perception (Fowler, 1985; Liberman, Cooper, Shankweiler, & Studert-Kennedy, 1967).

One example of this is the effect of the phonemes /s/ and /ʃ/ on the pronunciation and perception of neighboring stop consonants, particularly /d/, /t/, /g/, and /k/. There are two sorts of coarticulatory effects. First, the phoneme /ʃ/, the “sh” sound in *ship*, is produced with a rounding of the lips, which causes an elongation of the vocal tract that may persist through neighboring phonemes. Thus in saying “foolish dancer” the rounding of the lips extends from the /ʃ/ into the following /d/ and /æ/, thereby coloring the sound that is produced in pronouncing these phonemes. In contrast, the phoneme /s/ is produced by retracting the lips, thereby shortening the vocal tract; this retraction may persist into neighboring phonemes as well. Thus the /d/ in “fearless dancer” is produced with a shorter vocal tract than the /d/ in “foolish dancer.” The effect of this on the acoustic properties of the /d/ is to shift the distribution of acoustic energy into a lower energy range when the /d/ follows /ʃ/, compared with when it follows /s/. A second factor is that /s/ has an alveolar place of articulation, whereas /ʃ/ is a palatal sound. As a consequence, the front cavity is shorter in the case of /s/ than /ʃ/ and its spectrum is higher. The pronunciation of the alveolar /s/ may pull the place of articulation of

nearby velar and, perhaps, alveolar stops forward, causing them to have higher spectra than in other phonetic environments (Repp & Mann, 1980).

Listeners compensate for these effects by adjusting the boundary between phonetic categories which are distinguished by the frequency distribution of acoustic energy (Mann & Repp, 1981; Repp & Mann, 1981). This compensation can be demonstrated in the following way. It is possible to construct a sequence of sounds that ranges from a /t/ to a /k/, or from a /d/ to a /g/, by progressively lowering the distribution of acoustic energy contained in the noise burst which occurs on release of the consonant. This sequence of stimuli forms a graded continuum of sounds, and there is a boundary between the two percepts. The compensation for the coarticulatory influence of the /ʃ/ on the following phoneme takes the form of a shift in the perceptual boundary between the /t/ and /k/ sounds, as illustrated in Fig. 2. The figure shows that a sound that is ambiguous between /t/ and /k/ will be identified as a /t/ more often when following a /ʃ/, but as a /k/ more often when following a /s/.

In TRACE, we have assumed that these coarticulatory influences occur in the basic perceptual mechanisms that identify the phonemes of speech. This assumption is quite generally shared. Such influences are

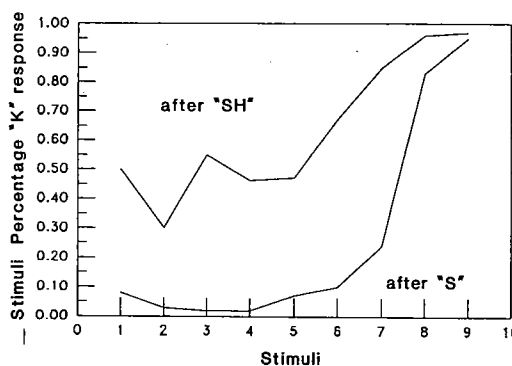


FIG. 2. Effects of a preceding fricative on identification of stimuli which range perceptually from /t/ to /k/. The graph at the left shows responses when the stimuli are preceded by /s/; the graph at the right shows responses when the stimuli are preceded by /ʃ/. (Figure adapted from Fig. 1 of Mann & Repp, 1981).

ubiquitous, and it has been suggested that one of the essential functions of the apparatus used for the perception of speech is to factor out the contextual influences and recover the underlying phonetic code (Fowler, 1985; Liberman et al., 1967).

The TRACE model accounts for these effects by allowing active phoneme units to modulate the connections between the feature level and phoneme level in adjacent time slices in just such a way as to compensate for their coarticulatory influences. That is, to account for the specific effect mentioned above, the units for /s/ and /ʃ/ would modify the strength of the connections between the feature units that feed into the phoneme units for /d/ and /g/, as well as /t/ and /k/, in earlier time slices to "undo" the coarticulatory effect that /s/ and /ʃ/ had on the stops. These modulatory connections provide the model with a powerful tool for adjusting perception in accord with context, and improve its performance relative to the case when the connections are disabled, as we have demonstrated in applications of the model to the perception of real speech (Elman & McClelland, 1986).

In the past, we have studied the behavior of TRACE by always testing the model under conditions where the compensation for coarticulation was triggered by the unambiguous presence of some phoneme which was known to have coarticulatory effects on adjacent sounds. However, in TRACE, phonemes receive input not only from the acoustic/phonetic level, but also from word units. Thus it is possible that activations at the phoneme level might trigger compensation for coarticulation whether they are purely determined by bottom-up input from the feature level or by top-down influences from the word to the phoneme level in addition to these bottom-up influences.

#### *Triggering Interlevel Effects via Intralevel Influences*

It is this possibility within TRACE for

top-down processing to have indirect effects on lower levels which suggests a way to answer the question: Are top-down effects real or only apparent? The logic is to test for the presence of a top-down effect not by seeing whether a higher level can alter an overt decision about a lower level, but, rather, by seeing whether the higher level can actually reach into the lower level and affect some intralevel process—in this case, compensation for coarticulation.

The following two simulations from TRACE illustrate the model's prediction that such effects will occur. In the first simulation, the model was given mock-speech stimuli consisting of the word "abolish" or "progress," followed by one of seven sounds which formed a continuum from an unambiguous /g/ at one extreme to an unambiguous /d/ at the other. The model located a perceptual boundary roughly in the middle; however, the precise location of this boundary varied depending on whether the preceding input was "abolish" or "progress." The boundary was shifted in accordance with TRACE's compensatory mechanism, as shown in Fig. 3a. In the second simulation, similar inputs were presented; however, the final fricative from both context words (the [ʃ] in "abolish" and the [s] in "progress") was deleted and replaced in both cases with an input which was exactly half-way between the two sounds. As can be seen in Fig. 3b, essentially the same boundary shift occurs. This is because the lexical information in the context word causes the model to perceive the ambiguous fricative in the appropriate way, so that the compensatory process occurs even though the bottom-up input has not specified the triggering phoneme.

It is important to note that the effect is weaker for perceptually restored phonemes than it is for phonemes that are actually present. The reason for this in the simulation is that both /s/ and /ʃ/ are partially activated when the input is ambiguous due to the balanced bottom-up support for each phoneme; the top-down effect strengthens

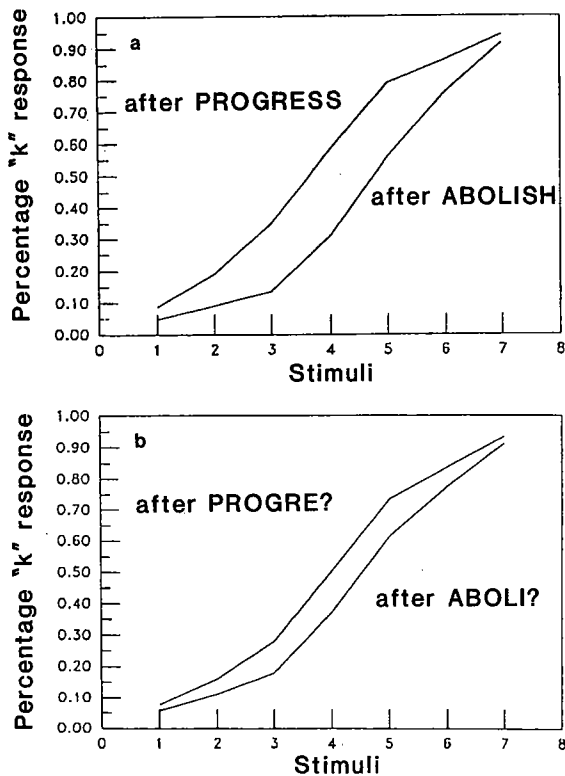


FIG. 3. Results of a computer simulation of the TRACE model. In a, the model identified sounds ranging from *t* to *k*, preceded either by "abolish" (left curve) or "progress" (right curve). In b, the model identified the same stimuli as shown in a; however, the final sound of the preceding context word was replaced by a sound intermediate between /s/ and /ʃ/ (represented as "?").

the activation of the contextually appropriate phoneme over the other, but this does not eliminate the activation of the other entirely. Compensation for coarticulation is graded in the model (see Elman and McClelland, 1986, for details), and so there is greater compensation when lexical and acoustic information are consistent.

In the following series of experiments we report the results of our test of the model's predictions. As with the simulations, the strategy was to test the status of the apparent top-down interaction by seeing whether it can trigger a second process which operates wholly within the lower (phoneme) level. If these lexical influences actually feed back information to the basic perceptual mechanisms that interpret speech sounds, then they ought to induce

coarticulatory compensation. On the other hand, if these lexical influences come only at a decision stage at which outputs of the perceptual mechanisms are interpreted, and do not feed back their results into the mechanisms responsible for coarticulatory compensation, we should not see coarticulatory compensation for phonemes whose identity is determined by lexical influences.

The critical point is that the task makes it possible to decouple the effect of a top-down interaction from the diagnostic used to detect it. In previous tests of top-down effects, lexical interactions have been measured through subjects' responses involving the target words themselves or phonemes contained within them. These responses might reflect either perceptual or postperceptual decisions, and so do not provide reliable evidence for the interaction. In the experiments reported here, the test of lexical interaction did not require a response involving the lexical item itself. Instead, the test was to see whether the lexical item could trigger a perceptual effect in an adjacent word. Because the response involved an unrelated word, and because the effect (compensation for coarticulation) is uncontroversially perceptual, we believe the test rules out a postperceptual decision interpretation. This was the logic for the following set of experiments.

#### EXPERIMENT 1

We had two goals for Experiment 1: First, we wanted to replicate the phenomenon initially reported by Mann and Repp (1981), namely, that listeners perceptually compensate for the coarticulatory effect of /ʃ/ or /s/ on a following sound. Second, we wanted to see whether the same effect could be obtained, but with stimuli in which the identity of the /ʃ/ or /s/ was determined by lexical influences. Accordingly, we constructed stimulus pairs in which a context word (e.g., "Spanish" or "ridiculous") was presented either intact or with its final sound replaced by the same intermediate sound, which was chosen to

be perceptually intermediate between /s/ and /ʃ/. From the TRACE model, we predicted that the word-final fricative in the context word would bias the perception of a subsequent ambiguous /t-/k/ or /d-/g/ discrimination, whether the identity of that fricative were determined either by lexical and acoustic influences (in the intact condition) or by lexical influences alone (in the replaced condition).

### Method

*Stimuli.* We began by preparing three sets of target stimuli for the /t-/k/ or /d-/g/ discriminations. One set of stimuli varied perceptually from "tapes" to "capes," another varied from "dates" to "gates," and a third from "deer" to "gear." The continua were created by first carrying out an acoustic analysis of naturally produced tokens of "tapes," "capes," "dates," "gates," "deer," and "gear." A computer program was used to edit these tokens in order to produce seven stimuli which varied in equal acoustic increments from /t/ to /k/ and /d/ to /g/. The parameters which were manipulated were (a) the frequency and amplitude of the noise burst following release of the stop consonant; (b) the onset frequency and trajectory of formant transitions; (c) the formant transition durations; and (d) the voice onset time of the stops, since this is known to covary with place of articulation. It is important to note that the step size—both acoustic and perceptual—between stimuli was deliberately made small. Even the first and last stimuli in each continuum were not unambiguous exemplars of their categories. In each case, each of the seven stimuli could be preceded either by an intact or by an altered version of a context word. For the intact conditions, the stimuli on the "tapes"–"capes" continuum were preceded by "Christmas" or "foolish"; the "dates"–"gates" stimuli were preceded by "copious" or "English"; and the "deer"–"gear" stimuli were preceded by "Spanish" or "ridiculous." For the replaced conditions, the

final sounds of these context words were replaced by sounds that were perceptually intermediate between /s/ and /ʃ/. These replacement sounds were constructed separately for each of the three continua by digitally excising the final /s/ and /ʃ/ from the original naturally produced tokens. These sounds were modified with a computer program to create a continuum of sounds ranging from /s/ to /ʃ/; then a separate group of subjects was asked to identify the sounds. That sound which was closest to being identified as /s/ 50% of the time was chosen as the replacement sound, to be used in place of the final sound from both of the context words used with each of the target continua.

Thus, instead of hearing, for example, "English [d/g]ates," or "copious [d/g]ates," subjects in the replaced condition heard "EngliX [d/g]ates" or "copiouX [d/g]ates," in which *both* the final sound of the context word (here shown with an X) and the initial sound of "dates"/"gates" were ambiguous.

*Procedure.* Three separate groups of 10 subjects were used in the experiment, with each group of subjects being tested on target stimuli from a different target continuum. Each subject participated in both the intact and the replaced versions of the experiment, with order of condition counterbalanced across subjects. In a given condition, a particular subject heard each of the seven stimuli on the target continuum 20 times, 10 times with one or the two context strings (e.g., "English" or "EngliX") and 10 times with the other (e.g., "copious" or "copiouX"). The target words followed the context words without any intervening pause, as though they had been pronounced together.

At the beginning of the experiment, subjects were told that they would hear a sequence of word pairs, and that they should try to identify the initial sound of the second word. Responses were indicated by pressing the appropriate key on the keyboard. The alternatives allowed were re-

stricted to the two endpoints of the appropriate continuum, either /t/ or /k/, or /d/ or /g/, as appropriate. Subjects were told that the first sound of the second word was often not pronounced clearly, and that they should do their best to indicate which of the two alternatives it sounded most like, even if they felt they were guessing. A response was required on every trial. The interval between trials was approximately 4 s. The instructions also indicated to the subjects that none of the word pairs made sense, and that they should not try to think of them as sensible phrases. The word pairs were described simply as pairs of unrelated words.

Before beginning the experiment proper, each subject received a sequence of practice trials with a different continuum than that used for the experiment proper for that subject. Following this practice, each subject received the 140 trials from either the intact or replaced condition; the other condition was run in a second session, on the next day or shortly thereafter.

**Subjects.** Thirty undergraduates at the University of California, San Diego, served as subjects. All were native speakers of English with no known speech or hearing disorders.

### Results

Our analysis of the results looks first to see if a reliable context effect occurred in each of the two conditions. Following this, we consider the relative size of the context effect in the intact and replaced conditions, and check for possible order effects.

The results for the intact conditions of Experiment 1 are shown separately for each target continuum in Fig. 4. Each panel of the figure displays two labeling functions. In one case, the percentage of /k/ (or /g/) responses is shown when the target stimuli were preceded by a word ending in /s/; in the second case, the target stimuli were preceded by /s/. As expected, the curves rise from left to right, reflecting the fact that the target stimuli vary along a /t/-

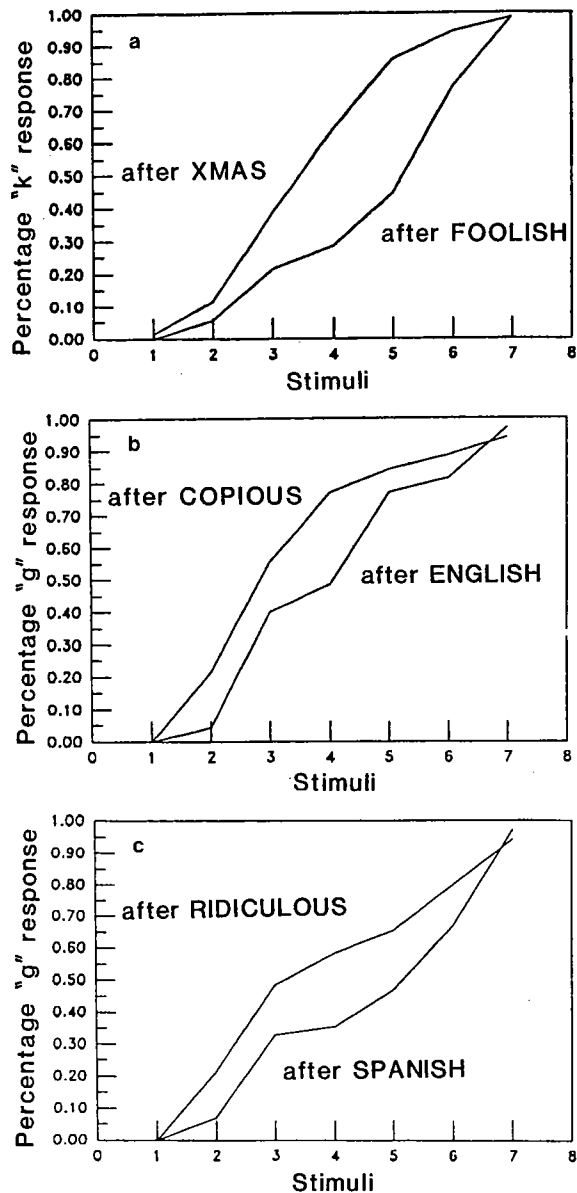


FIG. 4. Identification curves for three sets of experimental stimuli using intact context words. In a, stimuli ranging from "tapes" to "capes" were preceded by the word "Christmas" (left curve) or "foolish" (right curve); in b, stimuli ranging from "deer" to "gear" were preceded by the word "copious" (left curve) or "English" (right curve); in c, stimuli ranging from "dates" to "gates" were preceded by the word "ridiculous" (left curve) or "Spanish" (right curve).

/k/ or /d-/g/ continuum, as intended. The main effect of the target stimulus was highly significant in all three cases, with  $p < .001$  ( $F(6,54) = 82.881$ ) for "dates/gates,"  $p < .001$  ( $F(6,54) = 87.172$ ) for "deer/gear," and  $p < .001$  ( $F(6,54) =$

95.629) for "tapes/capes." Of central importance here is the fact that the preceding word affected perception of the following target sounds as follows: When a sound which is ambiguous between /t/ and /k/ (or /d/ and /g/) is immediately preceded by /š/, it is more likely to be perceived as a /t/ (or /d/). When the ambiguous sound is preceded by /s/, it is more likely to be heard as a /k/ (or /g/). The responses were subjected to an ANOVA, and the two conditions, target following /š/ vs following /s/, were significantly different, with  $p < .001$  ( $F(1,9) = 69.456$ ) for "dates/gates,"  $p < .001$  ( $F(1,9) = 26.938$ ), for "deer/gear," and  $p < .005$  ( $F(1,9) = 13.654$ ), for "tapes/capes."

Note that the result of this shift in boundaries is to undo the coarticulatory effect that /š/ and /s/ ordinarily have on the following stop consonant. Since /s/ makes a following /k/ sound more /t/-like, the perceptual mechanisms compensate for this, and move the boundary between /k/ and /t/. In our experiments, the stimuli on the /t/-/k/ continuum were originally produced in isolation; the shift in the boundary reflects what would have been an appropriate compensation for coarticulation, had the /t/-/k/ stimuli actually been produced following the /š/ or /s/ context.

Having validated our methods by verifying that listeners do indeed compensate for coarticulation, we are now in a position to ask whether or not this phonological compensation process can be triggered by information from the lexical level. This question can be addressed by considering the results obtained from the replaced conditions of the experiment. These results, shown graphically in Fig. 5, indicate that perceptual compensation does occur. As in Fig. 4, there are two identification functions for each of the three /t/-/k/ or /d/-/g/ continua. One curve shows the percentage of *k* (or *g*) identifications when the target is preceded by a context word which originally ended in /š/, and the second curve shows the responses when the target is pre-

ceded by a context word which originally ended in /s/.

The results are striking. Even though both context words end in a sound which is physically identical, listeners perceive the following sounds as if they were compensating for coarticulatory influences of the contextually appropriate *interpretation* of the preceding sound. An analysis of variance revealed that the two conditions (target following a context word which originally ended with /s/ vs /š/) were significantly different, with  $p < .001$  ( $F(1,9) = 42.882$ ) for "dates/gates,"  $p < .04$  ( $F(1,9) = 6.191$ ) for "deer/gear," and  $p < .02$  ( $F(1,9) = 8.442$ ) for "tapes/capes." Once again the main effect of the target stimulus was overwhelmingly reliable in all cases, with  $p < .001$  ( $F(6,54) = 97.407$ ) for "dates/gates,"  $p < .001$  ( $F(6,54) = 72.866$ ) for "deer/gear," and  $p < .001$  ( $F(6,54) = 129.932$ ) for "tapes/capes."

Additional analyses of variance were carried out for each continuum to determine whether the effect of context was significantly greater in the intact condition than in the replaced condition (as expected from the simulation results) and to determine whether there were any effects of order of the conditions on subjects' performance. Table 1 indicates that the effect was larger in the intact than the replaced condition for all three analyses. However, the Intact-Replaced  $\times$  Context word interaction was only significant for "dates/gates,"  $p = .038$  ( $F(1,9) = 5.878$ ), with  $p = .149$  ( $F(1,9) = 2.486$ ) for "deer/gear" and  $p = .312$  ( $F(1,9) = 1.145$ ) for "tapes/capes." There were no significant effects for order of presentation of the two conditions or significant interactions involving this factor.

### Discussion

Experiment 1 reveals that a compensatory adjustment in the phoneme boundary for a target phoneme did occur, based on the lexically determined identity of the preceding phoneme. These results suggest that

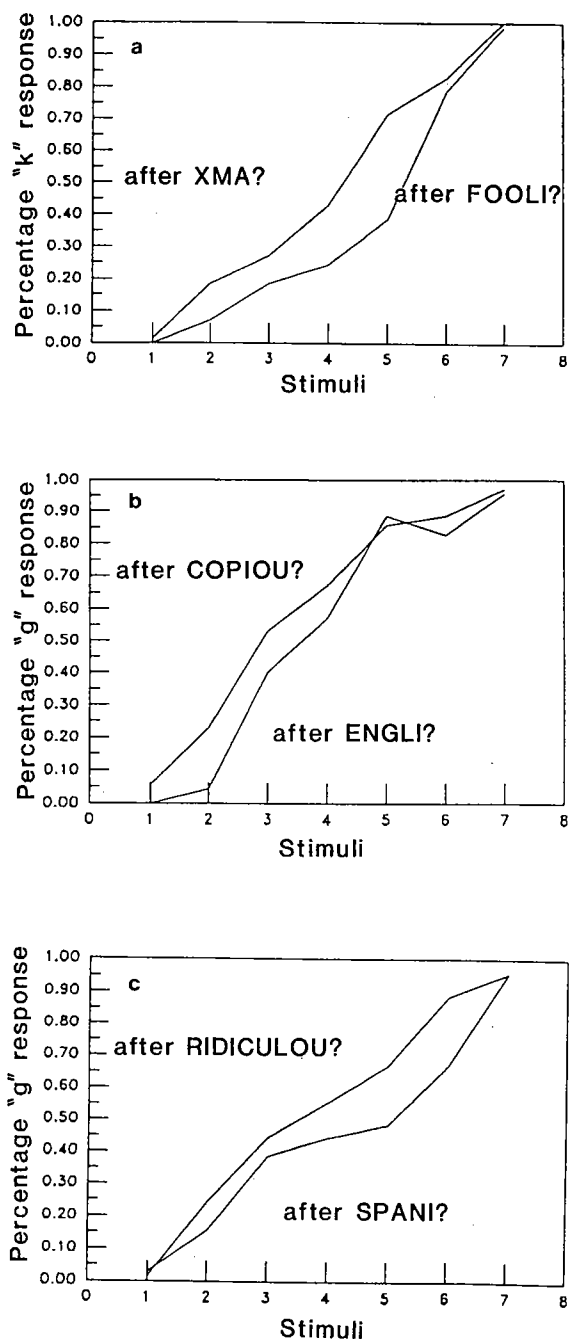


FIG. 5. Identification curves for three sets of experimental stimuli preceded by context stimuli in which the final sound, /s/ or /ʃ/, was replaced by a sound intermediate between /s/ and /ʃ/; this intermediate sound is represented as a "?". In a, stimuli ranging from "tapes" to "capes" were preceded by the word "Christma?" (left curve) or "fooli?" (right curve); in b, stimuli ranging from "deer" to "gear" were preceded by the word "copiou?" (left curve) or "Engli?" (right curve); in c, stimuli ranging from "dates" to "gates" were preceded by the word "ridiculou?" (left curve) or "Spani?" (right curve).

there are indeed top-down effects on phonemic processing. However, there are other possible explanations for these results which would be compatible with a bottom-up-only flow of information. One possibility is that the perceptual effect arose from acoustic information embedded in earlier portions of the context stimuli. This might have occurred because the vowel just preceding the ambiguous [ʃ/s] stimuli differed slightly in the two context words (e.g., "Spanish" vs "ridiculous"). Perhaps the compensation was for long-distance coarticulation between these vowels and the initial stops in the target words. Alternatively, the vowels themselves might not have triggered coarticulatory compensation in the /d-/g/ discrimination, but they may have contained coarticulatory cues to the identity to the following fricative which could have influenced the way in which that sound was perceived. These two alternative accounts share the prediction that the context leading up to the vowel should not be necessary to trigger the compensatory changes in the target discrimination. We tested this possibility in Experiment 2.

EXPERIMENT 2

This experiment examined the effects of the vowels used in the contexts associated

TABLE 1  
EFFECT OF CONTEXT ON PROBABILITY OF /g/ OR /k/ RESPONSES

	Context condition		Difference
	Ends in /s/	Ends in /ʃ/	
"dates-gates"			
Intact	0.53	0.41	0.12
Replaced	0.54	0.45	0.09
"deer-gear"			
Intact	0.60	0.50	0.10
Replaced	0.60	0.53	0.07
"tapes-capes"			
Intact	0.56	0.39	0.17
Replaced	0.49	0.38	0.11

with the "[d/g]ates" targets. In one condition, the context consisted simply of the vowels from the last syllables of the words "Spanish" and "ridiculous". In another, the vowels were presented together with the neutralized [s/š] fricative from the replaced condition of Experiment 1. In both conditions, the vowel is the only possible source of information that differs between the two contexts. If the vowel is indeed triggering the compensation in the /d/-/g/ discrimination, then the effect should show up in one or both of these conditions.

### Method

*Stimuli.* The target stimuli for both conditions of the experiment were the seven items varying from "dates" to "gates" that were used in Experiment 1. For the vowel-only condition, the context stimuli were the excised vowels from the "Spanish" and "ridiculous" context stimuli used in Experiment 1. The vowels were digitally excised from the stimuli used in Experiment 1 and consisted of that portion of the waveform in which none of the surrounding consonantal information could be perceived. Each vowel was preceded by a silent inter-trial interval of 4 s, and followed by a silent interval of 0.250 s, after which the target stimulus occurred. For the vowel-fricative condition, the same vowels were used, but each was followed by the neutralized [s/š] segment used in the replaced conditions with "Spanish" and "ridiculous" in Experiment 1. The CV context was followed immediately by the target stimulus.

*Subjects.* Eighteen subjects from the same sources as Experiment 1 were used in Experiment 2. None had participated in Experiment 1.

*Procedure.* Nine subjects were assigned to the vowel-only condition and nine were assigned to the vowel-fricative condition. Each group received practice with different stimuli of the same type as those used in the main experimental trials. The procedure was identical to Experiment 1, except

that (a) the context stimuli were different as described above; (b) the instructions were modified to reflect this difference; and (c) each subject was run in a single session only since each contributed data only to one condition.

### Results

Neither condition of Experiment 2 produced a reliable effect of context (see Fig. 6). In the vowel-alone condition, the excised vowels ([I] from "Spanish" and [ə] from "ridiculous") had no detectable effect on perception of the initial consonant in the target ( $p > .697$ ,  $F(1,8) = 0.163$ ). In the vowel-fricative condition, there was no reliable difference when these vowels were combined with the neutralized fricative used with "Spanish" and "ridiculous" in the replaced condition of Experiment 1 ( $p > .193$ ,  $F(1,8) = 2.022$ ). However, as Fig. 6 indicates, there may be a nonsignificant trend in the direction we would expect if the vowel-fricative pair were triggering coarticulatory compensation.

### Discussion

Although we found no reliable effect due to the excised vowel in Experiment 2, there was a trend in the vowel-fricative condition toward an effect of the excised vowel. Such an effect, if reliable, would have indicated that the vowel portion of the speech stream actually did contain information that influenced the perception of the subsequent /d/-/g/ segment. Since such information would be part of the acoustic input from the context stimuli used to induce the /d/-/g/ boundary adjustment in Experiment 1, it would represent a confounding of an acoustic cue with what was intended as a purely lexical influence. Although the effect of the vowel was not reliable in either condition, we felt that there was enough of a suggestion of such an effect that it would be important to try to find a way to isolate the lexical effect.

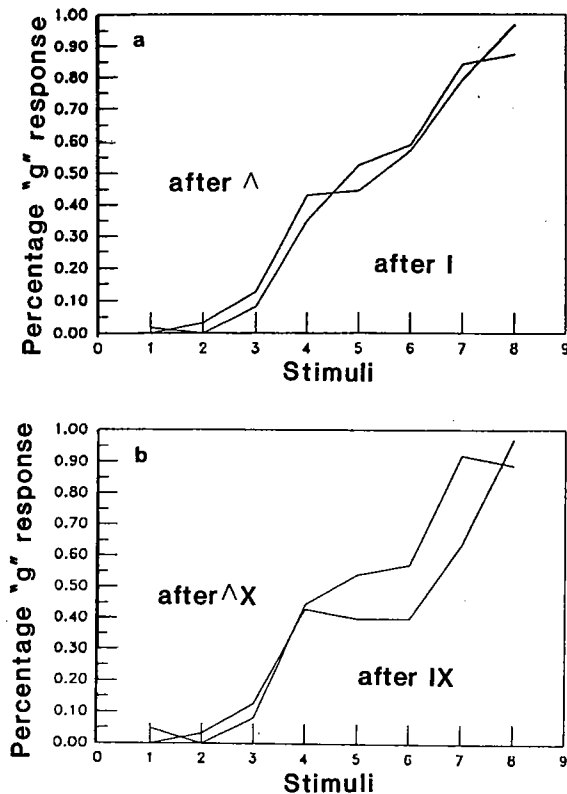


FIG. 6. Identification curves for "dates/gates" target stimuli when preceded in a by the final vowels in "Spanish" and "ridiculous" ([I] and [@], respectively); and in b when preceded by these same vowels plus a neutral fricative (represented by an X).

### EXPERIMENT 3

In this experiment, we devised two new sets of context materials, used in two new context conditions. In one of these conditions, called the VC-replaced condition, the final VC portions of the context words "Spanish" and "ridiculous" were replaced with an ambiguous version, pretested to sound halfway between [Iš] and [I@s], here designated [I^X]. In the other of these conditions, called the syllable-replaced condition, the final syllable of the context words "foolish" and "ridiculous" was replaced with an ambiguous version, pretested to sound halfway between [Iš] and [I@s]; this syllable is here designated as [I^X].

#### Method

**Stimuli.** The targets for both conditions

of the experiment were the seven "dates/gates" items used in Experiments 1 and 2. In the VC-replaced condition, these targets were preceded by "ridicul[X]" or "Span[X]". The neutralized [X] stimulus was selected from a set of candidate VC segments as the one judged to be closest to halfway between [Iš] and [I@s]. Thus, the final fricative and the vowel preceding it were both ambiguous and identical in both words. The pronunciation of the vowel in both these words in normal speech is actually virtually identical anyway, and so this manipulation resulted in a natural stimulus. This composite VC segment, designated [X], was then appended to the "Span—" and "ridicul—" stimuli, yielding the two contexts.

In the syllable-replaced condition, the words "foolish" and "ridiculous" were chosen to provide the two contexts. Digital recordings of these words were edited to delete the final syllables ("—lish" and "—lous," respectively). A composite syllable was created under computer control which was intermediate between these two final syllables and was, thus, also ambiguous as far as the final fricative. This composite syllable, designated [I^X], was then appended to the "foo—" and "ridicu—" stimuli, yielding the two contexts. These composite stimuli were not as natural sounding as the "Span[I^X]" and "ridicul[X]" stimuli, but they were still quite recognizable as the words from which they were derived.

**Subjects.** Twenty subjects from the same sources as before who had not participated in Experiment 1 or 2 were used in this experiment.

**Procedure.** Ten subjects were used in the VC-replaced condition and 10 more in the syllable-replaced condition. For each group, each of the two context stimuli appeared 10 times in random order before each of the seven target "[d/g]ates" stimuli used in previous experiments. Instructions were identical to those used in Experiment 1, and the procedure was the same except

that each subject was run in a single session.

### Results

For both conditions of the experiment, results were in accordance with the hypothesis that lexical factors can determine the perception of an ambiguous speech sound in a way that allows it to influence the perception of other ambiguous sounds. That is, the lexically determined identity of the vowel-fricative combination at the end of each context word appeared to produce a coarticulatory compensation. The results are shown graphically in Fig. 7. In the VC-

replaced condition, there was a significant lexical effect in that subjects tended to hear [g] after "ridicul[X]," and [d] after "Span[X]" ( $p < .018$ ,  $F(1,9) = 8.352$ ). Likewise, in the syllable-replaced condition, there was also a significant main effect for context ( $p = .030$ ,  $F(1,9) = 6.652$ ) with subjects hearing the ambiguous [d/g] sound as [g] more often following the "ridicu[l^X]" context than after "foo[l^X]". In both cases, there were no significant interactions.

### Discussion

The results of Experiment 3 appear to establish that lexical factors can indeed be responsible for perceptual compensation and support the conclusion that the perceptual effects are triggered by information from the lexical level. It is worth noting that the syllable-replaced condition makes it very hard to argue that the source of the compensation effect is sublexical. One might have proposed that simple phoneme-to-phoneme sequential constraints could be incorporated within the phonological level. Possibly, these sequential constraints are such that they would lead subjects to predict that the final phoneme in "Spanish" was an /š/ but the final phoneme in "ridiculous" was an /s/, quite apart from specific lexical factors; it may be that [nI-] is more often completed with [š], while [I?-] is more often completed with [s]. However, in the syllable-replaced condition, the context in the replaced items is actually the same for three phonemes before the final fricative; the vowel in "foo—," and the last vowel in "ridicu—" are the same vowel, though they may have slightly different acoustic realizations due to coarticulation, and the next two sounds in the two contexts are both acoustically and phonetically identical in the syllable-replaced stimuli. Thus, any differential prediction of the identity of the final fricative would have to be based on "f—" vs "ridic—," and thus would seem to be attributable to

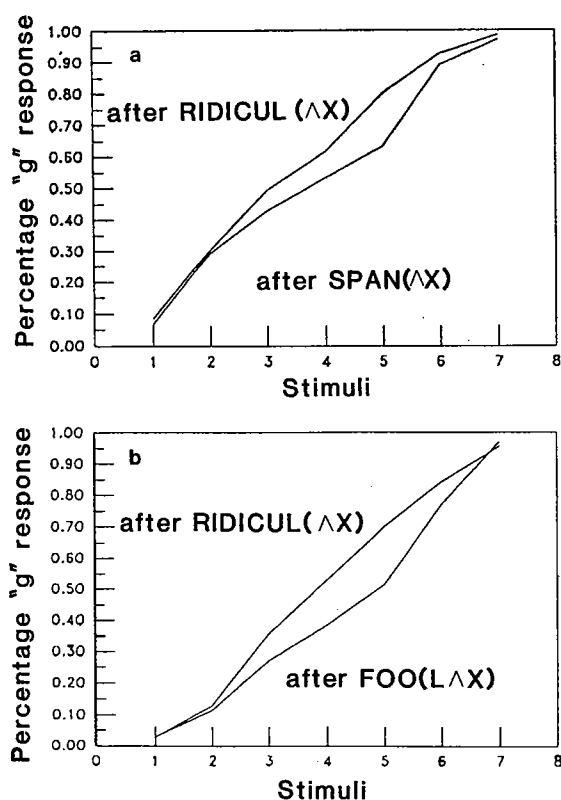


FIG. 7. Identification curves for ambiguous "dates"/"gates" stimuli. In a, stimuli were preceded by the word "ridicul[ $\wedge$ X]" (left curve) or "Span[ $\wedge$ X]" (right curve), in which the final [ $\wedge$ X] corresponded to the high central vowel [ɨ], followed by a sound intermediate between /s/ and /ʃ/. In b, stimuli were preceded by the word "ridicu[l $\wedge$ X]" (left curve) or "foo[l $\wedge$ X]" (right curve), in which the final [l $\wedge$ X] corresponds to a neutralized syllable beginning with [l], followed by the high central vowel [ɨ], followed by a sound intermediate between /s/ and /ʃ/.

knowledge that is specific to the particular lexical items involved.

The results thus far are compatible with our claim that the shifts we are seeing in the judgment of ambiguous /d-/g/ stimuli are due to coarticulatory compensation for the lexically determined perception of the final fricative in the context word. But one further alternative interpretation remains to be ruled out. Possibly, there is a subtle semantic bias in our materials that is influencing the results. Perhaps, for example, "Spanish gates" seems more plausible to subjects than "Spanish dates", or perhaps "ridiculous dates" seems more plausible than "ridiculous gates"; and perhaps these plausibility factors are determining subjects' /d-/g/ choices, rather than any coarticulatory adjustment to the word final fricative. Although the stimulus set was designed so that there would be no tendency for the context to actually allow subjects to *predict* the target words, it is nonetheless possible that there are subtle but significant differences in the plausibility of the different word pairs that influence subjects' decisions about the identity of the first sound in the target word. Experiment 4 was designed to examine this possibility.

#### EXPERIMENT 4

The two competing explanations for the tendency of subjects to hear sequences such as "Spani[s/š] [d/g]ates" and "ridiculous[s/š] [d/g]ates" as "Spanish gates" and "ridiculous dates" are as follows: One is that the lexical level affects perception of the ambiguous [s/š] sound so that listeners perceive it in the lexically appropriate way; and then this lexically biased percept triggers a compensatory perceptual effect in the following sound. The other explanation is that our particular choice of context and target stimuli favored interpretations in which, for semantic or pragmatic reasons, the word containing *d* or *g* (or *t* or *k*) was more appealing or plausible. If this were the case, then the apparent perceptual bias

simply reflects differences in semantic and pragmatic factors, and could be treated as a postperceptual phenomenon.

These explanations compete only in the case where the stimuli are presented auditorily. If the stimuli are presented visually, only what we will call the semantic factor is present; it is not likely that visual perception would be subject to compensatory perceptual mechanisms which owe their existence to facts about human articulation. To see if our effects were due to semantic factors, then, we presented visual analogs of the "dates/gates" target stimuli, preceded by visual versions of the two context words, "Spanish" and "ridiculous," used in several of the conditions described above. Two separate groups of 10 subjects were run with the same stimuli. One group was given instructions similar to those used in the auditory experiments, in which attention to the first word was down-played and the items were described as pairs of unrelated words. To give any possible effect of semantic and pragmatic factors the greatest chance of producing a reliable effect, the second group was given instructions which urged subjects to attend to both words and process the target stimuli as words rather than to focus just on the first letter of the target word.

#### *Method*

*Stimuli and display conditions.* Experiment 4 consisted of an analog to Experiment 1, except that stimulus presentation was visual rather than auditory. The stimuli were analogous to the "dates/gates" target stimuli and "ridiculous"/"Spanish" contexts used in several of the auditory conditions already described. Word pairs were displayed on an IBM PC monitor with a special low-persistence phosphor. The first word of each pair was either "SPANISH" or "RIDICULOUS" (in uppercase), and was presented centered on the display screen, following a 500-ms presentation of the word *READY*, centered at the same lo-

cation. The first word was displayed for 175 ms. This word was followed by a 75-ms interval during which an all-white mask was displayed by turning on all pixels in a centered row of 10 character blocks  $7 \times 14$  pixels each. This premask display was followed by a 25-ms presentation of the second target word, in lowercase, also centered in the display area. The target was followed by a 225-ms presentation of the all-white mask.<sup>3</sup> The second word was one of a set of seven stimuli which ranged visually from "dates" to "gates" (in lowercase). Each consisted of the letters *ates*, preceded by a specially constructed pattern of pixels, consisting of a lowercase *o*, and up to three additional pixels which tended to make the character appear more *d*-like or more *g*-like. The particular pixel patterns used were selected on the basis of pilot work to provide a set of closely packed items giving rise to about the same density of steps along a visual *d-g* continuum as the stimuli used in previous experiments did on an auditory continuum. Generally, successive stimuli along the continuum differed by a single pixel. Due to the pre- and post-masks, the impression of the second stimulus was rather faint, but the second stimulus was still visible as evidenced by the systematic shift from *d* to *g* responses across the seven stimuli, shown in the results section below.

*Subjects.* Twenty Carnegie-Mellon undergraduates served in the experiment to fulfill a course requirement. All were native speakers of English with normal or corrected-to-normal vision.

*Procedure.* The procedure was set up to be as analogous as possible to the proce-

<sup>3</sup> The all-white mask was intended to serve as a light mask, reducing contrast of the target. Though a contoured field would have produced stronger masking, we did not want to introduce possible bias effects of features from the mask on identification of the letters in the target word. The use of the all-white premask was intended also to buffer the target from any possible interactions of this kind with the context words.

dures used in Experiments 1 and 3. Each subject viewed 30 practice trials followed by 20 trials with each of the seven targets on the "dates-gates" continuum. Ten presentations of each target were preceded by RIDICULOUS and ten were preceded by SPANISH. One group of 10 subjects was told to focus on identifying the ambiguous letter at the beginning of the second word, and the response buttons were labeled simply *d* and *g* to encourage focusing of attention. These subjects were given the following instructions concerning the fact that the words occurred in pairs:

None of the word pairs really make sense, and you should not try to think of them as sensible phrases. They are simply pairs of unrelated words.

These instructions were taken word for word from the instructions used with the auditory stimuli in Experiments 1 and 3.

In order to try to give any subtle semantic effect a greater opportunity to occur, we altered these instructions for the second group of 10 subjects, stressing the importance of attending to both words and treating them as wholes. The choice alternatives were labeled "dates" and "gates" in an attempt to encourage subjects to treat the stimuli as whole words. These subjects were told that we wanted to determine whether the first word would influence their judgment, and that they must attend to both words for this to occur:

In this experiment, we are specifically interested in whether the first word of each pair influences your judgment of the identity of the second word of each pair. Thus, we want you to pay attention to both words of the pair. Your response should be based on your impression of which word was shown as the second word; accuracy is not as important as your impression, and your impressions are more likely to be affected by the first word if you attend to both words as wholes, so we would like you to make an effort to identify both the first and second words on each trial.

The wording in these instructions downplaying accuracy was inserted after two preliminary subjects reported that they felt

they had to try to ignore the first word in order to be as accurate as possible in reporting the identity of the first letter of the second word. Actually, accuracy is not always greater when subjects adopt a focused strategy (Jonston & McClelland, 1974).

### Results

There was no main effect of context on /d/-/g/ judgments in either condition of the experiment. The results are shown in Fig. 8. An analysis of variance of the results for the condition in which subjects were told to focus on the first letter in the second word indicates no significant main effect on perception of the ambiguous "[d/g]ates" word

as a function of the preceding word ( $p > .11$ ,  $F(1,9) = 2.977$ ) and a marginally significant interaction between the stimulus and context variables ( $p = .038$ ,  $F(6,54) = 2.418$ ). Even with explicit instructions to attend to both words, there was no main effect of context on the perception of the ambiguous letter in the second word. There was no main effect of context ( $p = .950$ ,  $F(1,9) = 0.003$ ), and no interaction between stimulus and context ( $p = .786$ ,  $F(6,54) = 0.526$ ). There appears to be a slight preference for perceiving *g* after SPANISH, which is opposite of the prediction for the auditory case; but this trend is reversed for the most *g*-like stimulus.

### Discussion

We did not find an effect of the visual context words "ridiculous" and "Spanish" on decisions about a visually ambiguous /d/-/g/ stimulus. What little tendency there is toward an effect of these words appears to be in the direction opposite that of the effect we obtained in the auditory versions of the experiment; if anything there is a tendency to perceive ambiguous *d-g* stimuli as more *g*-like following SPANISH, rather than more *d*-like as we found in the auditory case. Thus, Experiment 4 gives us no reason to suppose that the results we found in earlier experiments are the result of semantic or pragmatic biases in our materials. On the other hand, it must be conceded that Experiment 4 is not completely definitive. For one thing Experiment 4 rests on a null effect, and it is possible that a more sensitive method could be devised that would pick up a subtle bias effect. Another caveat is that we have tested for semantic/pragmatic effects using visual analogs of the stimuli used in the previous experiments. We see no reason to suppose that such an effect would be more potent with ambiguous auditory stimuli than with ambiguous visual stimuli, but the possibility remains that it is.

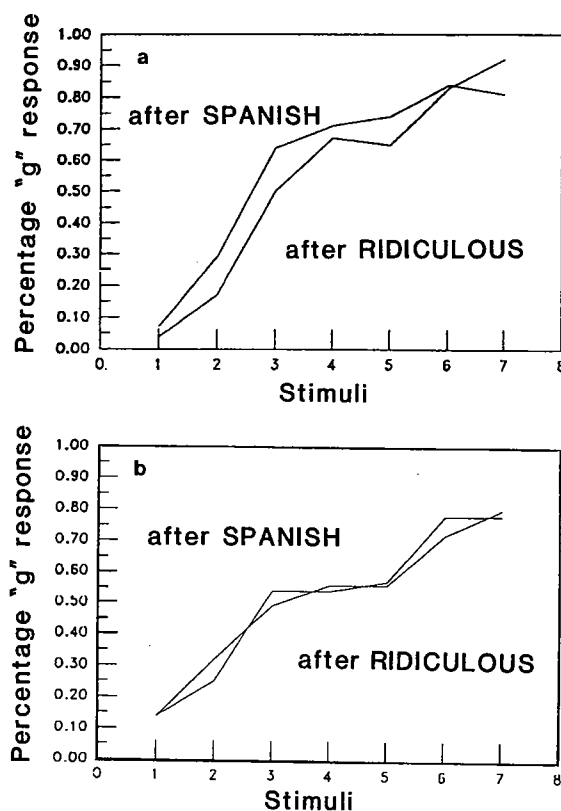


FIG. 8. Identification curves for ambiguous visual "dates/gates" stimuli, preceded by the context words RIDICULOUS and SPANISH, as indicated. In a, results are shown for subjects who were given instructions stressing attention to the first (target) letter of the second word. In b, results are for subjects who were given instructions stressing attention to both words as whole words.

Though it was only reliable in one case, there was a tendency toward an interaction of the context factor with the main effect of target stimulus. The most likely interpretation of the interaction is that RIDICULOUS, perhaps due to its greater length, reduced attention to the second word more than SPANISH did, thereby producing a flatter curve. This would make *g* more likely than *d* at the most *d*-like end of the continuum, and *d* more likely than *g* at the *g*-like end of the continuum.

### GENERAL DISCUSSION

Let us now ask once again what the basis is for the compensatory adjustment in the replaced conditions of Experiment 1. The sound preceding the ambiguous /t-/k/ and /d-/g/ stimuli was the same, equidistant between /ʃ/ and /s/. If compensation were based only on an analysis of the acoustic properties of the context, whatever coarticulatory effect this ambiguous context has should be the same in all cases. Experiment 3 rules out the possibility of covert acoustic cues in the vowel which might have triggered the adjustment. Apparently, then, what listeners are compensating for cannot solely be defined in terms of the acoustic stimulus.

In Experiment 4 we considered the possibility that the basis for the differing responses in the two contexts was not perceptual, but reflected higher-level knowledge about the relative likelihood of one word pair versus another. The results of this experiment provided no support for this possibility.

The most reasonable interpretation for our basic results is that listeners are compensating for phonemes whose identity is determined by lexical constraints. Although the final sound in each of the context words is ambiguous, the context words are such that only one of the possibilities is consistent with the word. Thus, "English" is a word, but "Englis" is not. The results

of this lexical influence appear to be made available to the parts of the perceptual mechanism that are responsible for compensating for coarticulatory influences on speech perception. This is indicated by the fact that the restored phonemes are able to trigger another perceptual phenomenon, which is the adjustment of phoneme boundaries to compensate for coarticulatory effects. It will be noted that this second phenomenon is clearly an influence of the identity of one phoneme, /s/ or /ʃ/, on another /t/ or /k/. It cannot reasonably be attributed to influences at the lexical level, since the /s/ or /ʃ/ sound is in a separate word from the /t/ or /k/. Since the phenomenon can be triggered by a lexically restored phoneme, this strongly suggests that the lexical restoration influences the mechanisms that compensate for coarticulation.

In motivating our experiments, we argued that if there were true top-down effects as assumed in the TRACE model, then we should expect to see compensation for coarticulation with lexically restored phonemes. But let us look at the question the other way around: Can the compensation for lexically restored phonemes that we have observed in our simulations be accounted for without postulating the existence of a true top-down effect? We think not.

Our argument hinges on what it means for some influence to be a true top-down effect. To us, it means that the influence goes back down and affects the same processing levels that information flows through bottom-up. Given this, our demonstration that lexical factors can trigger coarticulatory compensation is convincing evidence of a top-down effect, if it is accepted that the coarticulatory compensation actually occurs in the mechanisms that provide the phonemic information that serves as the basis for word recognition.

Is it true that the output of the mechanism that compensates for coarticulation does indeed serve as the input to word recognition? We think the answer to this ques-

tion is yes. If it were no, then compensation would influence phoneme identification, but would not influence word identification. This would lead to the paradoxical result that the identity of the first phoneme in the input "[t/k]apes" would reflect coarticulatory influences of previous phonemes but the identity of this same input as a word would not reflect these influences. More generally, it would lead to a nonsensical state of affairs in which lexical access did not benefit from the exquisite sensitivity of phoneme identification processes to local context.

The situation, then, would appear to be this: The mechanisms that provide input to word identification appear themselves to receive input from word identification. Influences run top-down, as well as bottom-up.<sup>4</sup>

As we have already indicated, our findings constitute just the sort of evidence that counts against the hypothesis that the mechanisms that perceive the individual speech sounds are "encapsulated" (to use the term introduced by Fodor, 1983, pp. 65ff) or isolated from the influence of higher levels.

Recently, Connine and Clifton (1987) have also described a very different line of experimental results that support the view that lexical information feeds back activation to the phoneme level. They contrasted the effects of a lexical manipulation with effects of a manipulation of payoffs. They found that both lexical factors and payoff manipulations produced beta-like effects on phoneme identification functions, but that lexical effects and payoff manipulations influenced reaction times in different ways; as predicted from an interactive account, lexical effects influenced reaction times to ambiguous stimuli, but not to unambiguous stimuli; in contrast, payoff ma-

nipulations influenced reaction times for unambiguous stimuli, not for ambiguous ones. Their interpretation was that the lexical manipulation influenced perceptual processes, while the payoff manipulation influenced the decision processes, as assumed in the TRACE model.

Neither our results, nor those of Connine and Clifton, rule out the possibility that levels of processing below the phoneme level are informationally encapsulated; just how far down higher-level factors can penetrate remains to be demonstrated through further experimentation. The TRACE model assumes that there are bidirectional connections throughout the multilevel processing system that underlies language perception and comprehension, and the model makes use of feedback from the phoneme to the model's feature level to account for the phenomenon of categorical perception, but we have not yet established conclusively that that feedback extends to the feature level itself.

It should be said, as well, that we have only established that *lexical* information can produce feedback into a lower level of processing. It remains to be seen whether constraints arising from even higher levels can have effects that percolate far enough down to trigger coarticulatory compensation. We hope that the technique of examining whether perceptual decisions whose outcome is influenced by higher level considerations can trigger compensatory adjustments will prove useful in testing for in-

---

overt phoneme identity responses that do not feed back into lexical access. Our results would be consistent with the possibility that only the latter of these mechanisms is able to make use of information about the identity of context phonemes that has been arrived at through influences from the lexical level. We have discounted this argument because it is completely unmotivated and unparsimonious, and it would still lead to the implausible conclusion that the lexically triggered compensation for coarticulation that occurred in our experiment influenced the identification of the first sound of the target word, but not the identification of the target word itself.

<sup>4</sup> It is logically possible that there are two separate mechanisms that are capable of compensating for coarticulation, one that is part of the main speech-processing stream and one that is used for generating

teractions among other levels. For example, one could ask whether semantic contextual influences on phoneme identity can trigger coarticulatory compensation, by using larger contexts in which the final context word is acoustically ambiguous (e.g., "mesh"—"mess"). Or one could ask whether such compensation could be triggered by an acoustically ambiguous word whose identity was constrained by the syntactic structure in which it occurs. Samuel's suggestion (1981) that sentence-level context does not reach down to the phoneme level would predict that such a top-down effect would not occur. Samuel's suggestion is consistent with the findings of Connine (1987); it would be useful to address this issue using the paradigm developed here.

Whatever the outcome of these future studies, we hope the present paper demonstrates the positive role a modeling approach can play in increasing the empirical base of psychological and cognitive science and in clarifying exactly which experimental tests actually bear on central issues of general importance beyond the specifics of the particular model under study. In this instance, thinking within the context of a specific simulation model made it clear why previous attempts to determine whether contextual influences really do feed back into the mechanisms of perception have not been conclusive and lead to an experiment that addresses, not only a specific prediction of this particular model, but also an issue that distinguishes interactive models as a class from noninteractive models. The specifics of the TRACE model are only one particular form in which this more general prediction of interactive models could have come to our attention. However, the model served a crucial heuristic role in that the prediction became apparent as a consequence of the configuration of assumptions built into the model.

More generally, this paper illustrates how modeling can uncover predictions that

depend upon the interplay of the assumptions that underlie the model and can lead us to ask new empirical questions. It will be the answers to these questions, together with ongoing attempts to build models that make sense of these answers, that ultimately lead to deeper understanding of the mechanisms of speech processing in particular and of cognitive processes in general.

#### REFERENCES

- COLE, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception and Psychophysics*, 13, 153-156.
- COLE, R. A., & JAKIMIK, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing* (pp. 149-211). London: Academic Press.
- COLE, R. A., & JAKIMIK, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-163). Hillsdale, NJ: Erlbaum.
- COLE, R., JAKIMIK, J., & COOPER, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America*, 64, 44-56.
- CONNINE, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, 26, 527-538.
- CONNINE, C. M., & CLIFTON, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 291-299.
- ELMAN, J. L., & MCCLELLAND, J. L. (1986). Exploiting lawful variability in the speech wave. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 360-385). Hillsdale, NJ: Erlbaum.
- FODOR, J. A. (1983). *The modularity of the mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- FORSTER, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.
- FORSTER, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology A*, 33, 465-495.
- FOWLER, C. A. (1985). *Segmentation of coarticulated speech in perception* (Status Report on Speech

- Research, Haskins Laboratories, SR-81, pp. 1-21).
- FOX, R. Unpublished manuscript. Vanderbilt University, 1982.
- GANONG, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- GROSJEAN, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28, 267-283.
- JOHNSTON, J. C., & MCCLELLAND, J. L. (1974). Perception of letters in words: Seek not and ye shall find. *Science*, 184, 1192-1194.
- JOHNSTON, J. C., & MCCLELLAND, J. L. (1980). Experimental tests of a hierarchical model of word identification. *Journal of Verbal Learning and Verbal Behavior*, 19, 503-524.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychology Review*, 84, 452-471.
- LUCE, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- MANN, V. A., & REPP, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 69, 548-558.
- MARSLÉN-WILSON, W. D. (1980). Speech understanding as a psychological process. In J. C. Simon (Ed.), *Spoken language generation and understanding* (pp. 39-67). New York: Reidel.
- MARSLÉN-WILSON, W., & TYLER, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71.
- MARSLÉN-WILSON, W. D., & WELSH, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.
- MASSARO, D. W. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 595-609.
- MASSARO, D. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language*, 27, 213-234.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 375-407.
- MORTON, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165-178.
- NORRIS, D. G. (1982). Autonomous processes in comprehension: A reply to Marslen-Wilson and Tyler. *Cognition*, 11, 97-101.
- REICHER, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 274-280.
- REPP, B. H., & LIBERMAN, A. M. (1984). *Phonetic categories are flexible* (Status Report on Speech Research, Haskins Laboratories, SR-77/78, pp. 31-53).
- REPP, B. H., & MANN, V. A. (1980). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, 69, 1154-1163.
- RUMELHART, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, NJ: Erlbaum.
- RUMELHART, D. E., & MCCLELLAND, J. L. (1981). An interactive activation model of context effects in letter perception: Part II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, in press.
- SAMUEL, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*.
- SAMUEL, A. G. (1986). The lexicon in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol 1. Speech perception*. Orlando, FL: Academic Press.
- TANENHAUS, M., CARLSON, G., & SEIDENBERG, M. (1984). Do listeners compute linguistic representation? In D. Dowty & A. Zwicky (Eds.), *Natural language parsing*. Cambridge, MA: Cambridge Univ. Press.
- THOMPSON, M. C., & MASSARO, D. W. (1973). Visual information and redundancy in reading. *Journal of Experimental Psychology*, 98, 49-54.
- WARREN, R. M., & SHERMAN, G. (1974). Phonemic restorations based on subsequent context. *Perception and Psychophysics*, 16, 150-156.

(Received August 4, 1987)

(Revision received November 9, 1987)