Simulating the time course of spoken word recognition: an analysis of lexical competition in TRACE

FRAUENFELDER, Ulrich Hans, PEETERS, Guus


Available at:
http://archive-ouverte.unige.ch/unige:83997

Disclaimer: layout of this document may differ from the published version.
**Simulating the Time Course of Spoken Word Recognition: An Analysis of Lexical Competition in TRACE**

Uli H. Frauenfelder  
*Université de Genève*  
Guus Peters  
*Max-Planck Institute for Psycholinguistics*

According to an emerging consensus in the psycholinguistic study of spoken word recognition, listeners first activate a set of lexical candidates and then select the appropriate lexical entry from among this set (Bard, 1990; P. A. Luce, 1986; Marslen-Wilson & Welsh, 1978). This view follows naturally from the properties of speech, in particular, its continuous and sequential nature. During most of the time that listeners are processing speech, they have only partial sensory information about any given target word—information that is insufficient to identify the word uniquely. They are, nonetheless, continuously generating or activating lexical hypotheses on the basis of this partial information. On the basis of this view of the word recognition process, the identification of any word form depends not just on the evidence in the input for the word itself, but also on the set of activated candidates or competitors partially supported by this evidence. Consequently, the time course of word recognition can be understood only with reference to the set of lexical competitors from which a given target word must be discriminated. For any target word and its competitor set, we are led to ask the following questions:

What are the lexical competitors of a given spoken target word at a given moment?  
How do these competitors influence the recognition of the target, and how does this influence vary as a function of the properties of the competitors individually and as a set?
How and when is the ultimate recognition choice for the target word over its competitors made?

Computer models have come to play an increasingly central role in the study of spoken word recognition because they allow researchers to address such questions directly. For example, the ability of one influential computational model, TRACE (McClelland & Elman, 1986), to generate exact predictions about the recognition time course of any word makes it particularly attractive. It is difficult, however, with such models to generalize from the simulated recognition performance on individual words to a deeper understanding of how different properties of words and the lexicon condition the model's behavior. Thus, it is still not clear what answers TRACE provides to these questions. Although claims have been made about the activated competitor sets in this model on the basis of intuitions or simulations with individual words, these have never been tested systematically, in part because most TRACE simulations have been concerned with another processing level, phoneme recognition and the influence of lexical feedback on speech processing (Elman & McClelland, 1988; Massaro, 1989; McClelland, 1991).

The goal of the simulations presented here is to study TRACE's predictions about the nature of the activated competitor set and the time course of word recognition. Such detailed simulation studies with TRACE are important because they allow us to evaluate one major class of solutions to the problem of lexical competition, that relying on lateral inhibition and top-down facilitation. Lateral inhibition at the lexical level in TRACE is assumed to carry the essential burden of reducing the many potentially activated lexical candidates and of converging on the target lexical entry.

Before presenting a series of TRACE simulations that aim to make explicit the answers to these questions, we briefly present four other models of spoken word recognition. These models represent different points in the evolutionary process from verbal to computer-implemented models. This analysis illustrates the changes in the way that models deal with the complexity faced in characterizing lexical processing. We compare their assumptions about the mechanisms of lexical competition and the resulting activated candidate set. We also consider the experimental results mentioned in the literature; these results must be taken into account when evaluating the competition mechanisms assumed in the different models.

**FOUR MODELS OF SPOKEN WORD RECOGNITION**

The verbal model Cohort I (Marslen-Wilson & Welsh, 1978) makes several simplifying assumptions about word recognition and produces several testable predictions about the time course of word recognition. Cohort II (Marslen-Wilson, 1987), also a verbal model, includes more complex and realistic processing assumptions than does Cohort I. Because Cohort II is not computer implemented, it is difficult to assess exactly what predictions it makes. In contrast to a verbal model like the Cohort II, the TRACE model (McClelland & Elman, 1986) is computer implemented and has the advantage of being able to make precise quantitative predictions about the time course of spoken word recognition. TRACE also has major limitations, some of which are addressed in a more recent computational model, Shortlist (Norris, 1994).

**The Cohort Model: Version I**

The original Cohort model (Marslen-Wilson & Welsh, 1978) represents the first attempt at providing systematic answers to the three previous questions. In doing so, this model assumes two successive stages of processing. During the first stage, all words that exactly match the onset (i.e., the initial one or two segments) of the target word are activated. In this way, a set of competitors or the word initial cohort of the target is generated. No other words are allowed to enter the cohort and to compete for recognition. After the initial activation phase is a stage of deactivation during which the cohort members that mismatch later arriving sensory input are eliminated from the cohort via bottom-up inhibition. The number of cohort members decreases as more stimulus information becomes available. Thus, in response to the question about the lexical competitors of a target word, the cohort model asserts that the competitors at any given time are those words that match and are aligned with the target word. Neither words that are aligned but that mismatch the target (even in only a single distinctive phonological feature) nor words that match the target but are not initially aligned count as competitors.

The answer to the question about the competitors' influence on the target is also simple. The mere presence of competitors in the cohort prevents the target from being recognized, but neither the number (cohort size) nor the nature (e.g., competitor frequency) of the cohort members affects the time course of word recognition in this model. Only the final cohort member(s)—the word(s) matching the target the longest—determine the target's recognition point.

Finally, the cohort model answers the question about recognition choice; it makes clear predictions about the moment that any word can be recognized in a given lexicon on the basis of an analysis of its cohort members. This recognition point (RP) is assumed to correspond to the word's uniqueness point (UP) or the moment that the word becomes unique with respect to the other words in the lexicon. A target word spoken in isolation is thought to be recognized when it is the only word remaining in its cohort. For example, a word like elephant, heard in isolation, is predicted to be identified...
at the sound /f/ at which point no other words in the lexicon share its initial part.

By its clarity and simplicity, Cohort I generates precise predictions about the time course of recognition of all words. The recognition of a word should be a direct function of the position of its UP. Although these predictions can be tested and falsified, some critics were quick to point out ways in which this attractive model fails to account for, among other things, the robustness of human language perception.

The Cohort Model: Version II

To incorporate some assumptions psychologically more realistic than those of Cohort I, Marslen-Wilson (1987) proposed a new version of this model, Cohort II. Like its predecessor, the new version assumes that the competitor set is limited to those candidates that are aligned with the onset of the target word. The membership of the cohort is enlarged to include words that mismatch the sensory input to some minimal but still unspecified degree (i.e., the input sequence nobility might activate mobility). To express the varying degree of match between the input and the different competitors, the model appeals to the notion of levels of activation. In particular, cohorts vary in activation level as a function of their fit with the input and their frequency. Unlike the clear binary status of words in the original model (either in or out of the cohort), the identity and status of cohort members is less well defined in this newer formulation of the model. The model does not specify how the degree of match with the input and how word frequency determine activation, because these factors and their relative contributions to lexical activation are difficult to quantify in a verbal model. As a consequence, no precise definition of the competitor set is available in this model.

As in the original version, the cohort members exert no direct influence on the target word or its activation level, but the decision about when a particular word is recognized does depend on the activation level of its competitors. Marslen-Wilson (1990) has suggested that isolated word recognition takes place when the difference in activation between the target word and its most activated competitor has reached a certain criterion. In this way, properties of cohort members, like their frequency, influence the decision phase of the recognition process. Because so many factors contribute to the recognition decision, it is impossible for this more recent version of the Cohort model to give a precise answer to the question about making a recognition choice.

This discussion of the two versions of the Cohort model illustrates a general dilemma confronting the effort to model lexical processing. Cohort I makes clear and testable predictions and therefore has been useful in generating considerable empirical research, but to do so, it has made several simplifying assumptions. In contrast, Cohort II is a more complex verbal model and presumably fits better with what we know about lexical processing, but it does not provide direct answers to the questions about the competitor set and therefore cannot predict the time course of word recognition.

The TRACE Model

TRACE is an interactive activation model that has been implemented as a computer program (McClelland & Elman, 1986). It is made up of distinctive feature, phoneme, and word units that each represent hypotheses about the sensory input. These three types of units are organized hierarchically. There are bottom-up and top-down facilitatory connections between units on adjacent levels (feature-phoneme, phoneme-word, and word-phoneme) and inhibitory connections between units in levels (feature-feature, phoneme-phoneme, and word-word). Incoming sensory input provides bottom-up excitation of distinctive feature units, which in turn excite phoneme units. Phoneme units are activated as a function of their match with the activated distinctive features so that several alternative phonemic units are activated for a given input. As the phonemes become excited, they increase the level of activation of words that contain them. As words receive some activation, they begin to inhibit each other. Finally, words also excite the phonemes that they contain in a top-down fashion. Figure 4.1 presents a schematic representation of the model and its connectivity pattern.

![Image of the TRACE model](image.png)

**FIG. 4.1.** Schematic representation of the connectivity pattern and of successive time slices in the TRACE model. From Frauenfelder, 1996. Reprinted with permission.
Time is represented spatially in TRACE. For each time slice, this model constructs a complete network in which all the units at every level are represented. Thus, to recognize an input made up of four phonemes, TRACE constructs at least four (in fact, four times six slices per phoneme) complete lexical networks and marks the time cycle at which each lexical unit begins. This solution of spatial reduplication is not particularly realistic or efficient. The model cannot easily account for the well-known psycholinguistic phenomenon of repetition priming because the same word is represented by different units situated at different time slices with no connections between them. Moreover, to recognize a sequence of words, the model must first wire the connections between and in the temporally distinct time slices. This solution is very costly from a computational point of view.

Compared with the Cohort model, TRACE generates a potentially much larger competitor set. Unlike both versions of the Cohort model, TRACE assumes exhaustive alignment (Frauenfelder & Peeters, 1990): The activated competitor set for a particular stretch of speech is not restricted to those lexical candidates that are aligned with particular points of alignment (i.e., like word onsets as for the Cohort model). Rather, the competitor set can include words matching the input but not aligned with word onsets. As a result, all words in the lexicon are in constant potential contention for recognition—continuously increasing or decreasing in activation level—at every processing moment as a function of the incoming signal.

TRACE diverges from both versions of the Cohort model in assuming that lexical competitors can exert a direct influence on the activation level of the target and vice versa. The model incorporates a lateral flow of inhibition between units at the same level. By this mechanism, the target word inhibits its competitors but is also inhibited by them. The degree to which one word inhibits another is a function of the former's activation level: The more a word is activated, the more efficiently it can inhibit its competitors. The dynamics of interactive activation and, in particular, of this inhibition between competitors allows TRACE to keep the actual activated competitor set small and to converge on a single lexical entry despite the mass of lexical candidates that are potentially in contention for recognition.

Another aspect in which TRACE deviates from both versions of the Cohort model is its assumption of a top-down activation flow from the word to the phoneme level. As a result, the members of the activated competitor set can exert an additional indirect influence on one another. The activated words provide top-down excitatory feedback to the phoneme units they contain by increasing the latter's level of activation. These phoneme units in turn again excite the connected word units. This "resonance" was first identified in the interactive activation model (IAM) for visual word recognition (Rumelhart & McClelland, 1982). Because this phenomenon has not been studied systematically in the auditory domain, it is the object of one of the simulations presented here.

### The Shortlist Model

The Shortlist model (Norris, 1994) represents an attempt to improve on some shortcomings of TRACE, in particular, its rather implausible architecture. Shortlist involves two distinct processing stages. During the first, a restricted set of lexical candidates, the shortlist, is established using both bottom-up excitation and bottom-up inhibition. Any word—whatever its alignment—can be included in the activated candidate set as long as it matches the input to some preset criterion. Unlike TRACE, where all the words are prewired into such a network, the modeler sets an upper limit (e.g., ranging from 3 to 30 words) on the number of candidates retained. For practical reasons, this shortlist is based on the input's match with the phonological form representations of words found in a lexical database. During the second stage, the best fitting words—even those with different alignments—are wired into an interactive activation network similar to that assumed by TRACE. The competition between the lexical items in the shortlist takes place via lateral inhibition, such that the words that best match the input inhibit their competitors most effectively.

The Shortlist model (Norris, McQueen, & Cutler, 1995) has been modified in two important respects. First, a mechanism was introduced in the model to exploit prosodic information and more specifically to incorporate the metrical segmentation strategy (MSS) developed for the segmentation of English (Cutler & Norris, 1988). To do so, the new model boosted the level of activation of words that have a strong initial syllable (not containing a reduced vowel) and decreased the activation of words that are not aligned in their onsets with a strong syllable in the input. Second, the model introduced a recomputation routine so as to prevent highly activated competitors from completely inhibiting later arriving matching words. By recomputing the activation of the shortlist after each phoneme is received, the model avoids committing itself completely to the lexical candidates that are activated early, and it allows later matching words to become activated.

### COMPARISON OF MODELS

We have now examined four major models of spoken word recognition. Other computational models of interest have been proposed, including those appealing to recurrent networks (Content & Sterenon, 1994; Elman, 1990; Norris, 1990. For a discussion of the relative advantages of localist over distributed models, see chap. 1, this volume). We have restricted ourselves
here to these models because of their relative homogeneity and widespread use. Moreover, they illustrate nicely the progression from verbal to computational models.

From Verbal to Computational Models

Simple verbal models are helpful in shaping and formalizing questions and issues, but psychological phenomena are generally affected by many variables, and language performance is no exception. Some of the factors that must be dealt with in word recognition research include the form properties of words (quality of sensory input, length, phonological structure, etc.), the grammatical and abstract properties of words (syntactic form class, semantic category, word frequency, and morphological structure), and the properties of the lexicon (number of competitors, form properties of competitors, grammatical and abstract properties of competitors). We cannot characterize most of these factors adequately in dichotomous categories as simple verbal models require. In most instances, psycholinguists are confronted with intricate interactions and the covariation of multiple factors.

Because verbal models are intrinsically limited in their ability to describe the influence of multiple factors and their interactions, they lead to inappropriate simplification to make predictions. This simplification takes at least two forms: limiting the number of factors taken into account and treating the factors as dichotomous rather than multivalued. One way of dealing with this problem is to implement models on a computer. As we intend to show in the following simulations, however, new challenges arise in interpreting the simulation results and in relating these results to the postulated underlying mechanisms.

Predictions About Competitor Set and Lexical Competition

It is important to directly compare the claims of the four models under scrutiny about the selection process and the resulting membership of the activated competitor set. The makeup of this set for any given model depends on its assumptions about three main factors: the nature of the input units used to contact the lexicon (e.g., phonemes or syllables), the manner in which this representation is aligned with the lexical representation, and the nature of the activation flow (bottom-up, top-down, and lateral). We focus here on the problem of activation flow. (For a discussion of the other factors, see Frauenfelder, 1996.) Table 4.1 summarizes the patterns of activation flow characterizing the selection mechanisms in these models.

These models appeal to different mechanisms to activate the competitors and subsequently to reduce the competitor set. According to the Cohort model, lexical selection is achieved by means of bottom-up activation combined with bottom-up inhibition. When mismatching sensory information is received, it deactivates the inappropriate lexical units. According to TRACE, competitor set reduction is achieved essentially through lateral inhibition. This inhibition between lexical competitors allows the stronger candidates, and in particular, the target, to dominate and eliminate the weaker ones. Bottom-up inhibition and lateral inhibition can also be combined as demonstrated by the Shortlist model.

These two mechanisms for activating and reducing the lexical competitor set can lead to divergent empirical predictions. Frauenfelder (1996) has examined in detail which lexical competitors are assumed to be activated for specific sensory inputs by each of the four models. A complete discussion of these contrasting predictions goes beyond the scope of this study, but by all accounts the amount of bottom-up activation received by any given word unit—target or competitor—represents this unit's fit with the sensory input. This fit depends on the quality and quantity of the match between the sensory input and the lexical form representation. Thus, all the models assume that cohort competitors—which match and are aligned with the input—are activated.

In contrast, the models make divergent predictions about the activation (or the nonactivation) of mismatching and misaligned competitors. First, the models' tolerance of mismatch in the sensory input varies. It can range from extreme intolerance and total lexical deactivation by any mismatch in the sensory input via strong bottom-up inhibition (as in Cohort I) to relative tolerance and partial activation despite some mismatch. In this case, some function presumably relates the phonological distance between the input and a given competitor to the level of activation attained by that competitor. The closer and more complete the fit, the greater the activation received by the word.

An obvious but crucial difference between the two selection mechanisms is that for models with lateral inhibition, the competitors have a direct influence on target activation and vice versa. Consequently, the quality of the input's fit
not only with the target but also with its competitors, the number of competitors, and their frequency all affect the level of relative activation of the words in the competitor set. For models without lateral inhibition, none of these factors except the fit with the target influences word recognition.

The following two sections review a selection of the experimental studies that have addressed the questions raised earlier about the activation of different types of competitors by word and nonword input sequences and the effects that these activated competitors can have on the recognition of the target word.

**Experimental Evidence on Competitor Activation**

There is at present considerable evidence for the multiple activation of lexical candidates, but it is impossible to treat all the different types of competitors in a homogeneous fashion and to make generalizations about their behavior. In fact, because of the variability in the predictions and the experimental results, it is necessary to examine the specific types of competitors individually as we do in this section.

**Activation of Matching and Aligned or Cohort Competitors**

There is clear experimental evidence that cohort competitors are activated during processing. One important study (Zwitserlood, 1989) using the cross-modal semantic priming procedure\(^1\) (Swinney, 1979) showed that the meanings of a target word and its cohort competitors are activated as long as the presented sensory information did not distinguish between them. Thus, with the partial spoken input [kept], the meanings of both captain and captive were activated. This finding is consistent with all four models presented here.

It is also of interest to consider what support the psycholinguistic literature provides for the notion of UP as this notion is based on cohort competitors. The relation between the location of the UP and the moment at which words are recognized has been investigated with different experimental procedures.

\(^1\)This procedure involves presenting subjects with a spoken sentence or a list of isolated spoken words (or nonwords). During or just after this auditory presentation, subjects receive a visually presented word for which they are required to make a lexical decision. In the critical conditions, the visual word is semantically related either to the auditorily presented word or to one of its competitors. The lexical decision latencies in these conditions are compared with the reaction times (RTs) obtained in matched unrelated control conditions. Any RT advantage in the lexical decisions in the related over the unrelated conditions is taken as evidence for the semantic activation of the auditorily presented word or its competitor. The moment during the presentation of the spoken word (or part thereof) at which the written word appears on the computer screen determines when lexical activation of the former is being measured. Generally, the visual word appears immediately after the offset of the entire spoken word or of a fragment of this word.

**Activation and Deactivation of Mismatching Competitors**

It is useful—especially for isolated word recognition studies—to distinguish mismatching sensory information that is located at item onset from that arriving later in the speech sequence. In the first case, we must ask whether mismatching competitors are activated at all, and if so, in which conditions. In the second case, we must determine what happens to the matching and hence activated competitors once a mismatching input does arrive. We consider each of these cases in turn.

**Onset Mismatch.** The results about the activation of initially mismatching lexical candidates are not yet completely clear. A first cross-modal semantic priming study in Dutch (Marslen-Wilson & Zwitserlood, 1989) pro-

\(^2\)The RTs' differences corresponded to only 30% or 50% of the measured physical distances between the UPs of the two conditions. Although this pattern suggests that the UP does not correspond exactly to the RP, other possible interpretations of this finding relate to the difficulty in determining the appropriate definition of the UP (which lexical representations\(^3\) etc.).
duced results suggesting that only matching lexical candidates are activated. These authors showed that a spoken word or nonword prime that mismatched the target word in its initial phoneme by several distinctive features did not activate the meaning of the target word (e.g., neither the rhyming word *mat* nor the rhyming nonword *dat* activated the semantic associate *dog* of the source word *cat*). This finding indicates that words differing by several distinctive features in their first phoneme do not belong to one another's cohort. In a follow-up study (Marslen-Wilson, Moss, & van Halen, 1996) with the same technique, some activation of mismatching competitors was obtained in certain conditions. More specifically, when the priming nonword was phonologically close to the source word and had no other rhyming competitors, significant priming was obtained, but less than when the source word was the prime. A second experiment produced conflicting results with no priming for either close rhyming words or nonwords. Connine, Blasko, and Titone (1993) also used the cross-modal priming technique and manipulated the phonological distance between a priming nonword and the target word. The results showed priming by nonwords that were created from words by making small phonological changes in their onsets. When this phonological distance was increased, no activation was obtained.

We can draw two main conclusions on the basis of these studies. First, lexical candidates appear to be activated, but only when they mismatch the sensory input initially by a small phonological distance. Second, the lexical status of the sensory input—word or nonword—affects the amount of lexical activation of mismatching competitors. Evidence for activation of lexical competitors was restricted to cases where the input was a nonword. These results appear to be more consistent with the predictions of the TRACE and Shortlist models than with those of Cohort I, which allows no activation of mismatching lexical competitors. In principle, the former two predict competitor activation, especially for nonword inputs. Indeed for these inputs, no exactly matching lexical candidate exists that would strongly inhibit any mismatching words.

A word of caution is in order here about the predictions of these models. First, the claim that TRACE activates competitors with initial mismatches and that this activation depends on the phonological distance defining the mismatch has never been tested directly. This definition is one of the objectives of the present paper. Second, although, in spirit, phonological distance determines lexical activation in the Shortlist, the current implementations of the model do not yet include this property because they are based on phonemes and do not take into account phonological distance.

**Noninitial Mismatch.** What happens when the mismatch comes later in the input and several lexical competitors have already been partially activated by matching information? Zwitserlood (1989) conducted some relevant experiments showing that activated candidates were immediately deactivated when mismatching information was received. When longer auditory primes that included some mismatching sensory information were presented (e.g., the *I* in *kreptl*), the semantic associate of the then mismatching competitor (e.g., *captain*) was no longer primed, and this competitor was assumed to be deactivated. Unfortunately, it is impossible in this study to determine whether the mismatching input eliminated this competitor in a bottom-up fashion or whether the matching competitor (e.g., *captain*) had received sufficient activation from the additional input to strongly inhibit this mismatching competitor. This confound was avoided in a phoneme monitoring study (Frauenfelder, Content, & Scholten, 1998) in which phonemic mismatches were introduced noninitially. This mismatching information */n/* preceded the phoneme target */r/* in the nonword (*vocabulaire*). The monitoring latencies did not differ from those in a control nonword condition in which no lexical activation is expected (*sotobinaire*), but were much slower than those in the original word (*vocabulaire*). This pattern of results suggests that the mismatching phoneme immediately deactivated the lexical candidate. These results are consistent with Shortlist, which assumes bottom-up inhibition, but inconsistent with TRACE, which includes only lateral inhibition.

**Activation of Misaligned (Embedded) Competitors**

Another class of potential competitors that must also be considered are words embedded somewhere in longer words. For example, during the processing of the word *party*, the embedded words *par* and *tea* could potentially be activated because they both match the input. It is important, in particular for models assuming lateral inhibition, to distinguish between both the position of the embedded word (initial and noninitial) and the lexical status of the embedding item (word or nonword). In TRACE, initially embedded words are activated in both longer word and nonword strings, whereas noninitially embedded words are activated only in nonwords (Frauenfelder & Peeters, 1990).

**Initially Embedded Words.** In a phoneme monitoring study, Frauenfelder and Henstra (1989) tested for the activation of embedded words by comparing the detection latencies to phoneme targets (e.g., for target */p/* in matched monosyllabic words (e.g., *map*) and nonwords (e.g., *nop*) in three conditions: baseline (e.g., *nomp* vs. *nomp*), initially embedded (e.g., *mopel* vs. *nelp*), and finally embedded (e.g., *tenap* vs. *tenop*). Because the acoustic properties of these targets and their local environment varied minimally across the pairs being compared, the faster RTs obtained for targets in embedded words than in embedded nonwords can be interpreted in terms of
the lexical activation in the former condition. Such an advantage occurred for both initially and finally embedded words, a finding suggesting that these embedded words are activated during the processing of the longer nonword string. These results are consistent with both TRACE and Shortlist.

**Noninitially Embedded Words.** Shillcock (1990) examined whether words, embedded noninitially in longer words, are activated, again with the cross-modal semantic priming task. He obtained significant priming of a word (e.g., *rib*) by the second syllable (e.g., *bone*) of a bisyllabic target word (e.g., *trombone*). This finding suggests that the embedded word is activated during the processing of the longer carrier word, a result consistent with the predictions of the Shortlist model but not of the TRACE model.3 This finding was not replicated by Gow and Gordon (1995), who obtained no cross-modal priming of a noninitially embedded word (e.g., *lips*) when presented with longer carrier words (e.g., *tulips*). It is at present not understood why these two studies produced contradictory results.

On the basis of these rather limited studies of the activation of embedded words, it appears that such words are activated in both initial and noninitial positions in word and nonword carrier sequences. The TRACE model, it should be recalled, predicts effects of the lexical status of the carrier sequence on embedded word activation. When the carrier sequence is a nonword and when no matching lexical candidates are aligned to the left of this target, the embedded word is activated. In the case of word carriers, noninitial targets suffer too much lateral inhibition from the carrier word to be activated. These predictions differ from those made by Shortlist, which, because of its reset mechanism, allows the activation of noninitially embedded words in longer words.

**Experimental Evidence on Lexical Competition**

In the preceding section, we examined the experimental literature on the activation of lexical candidates. Here we pursue the question of what effect these activated lexical candidates have on target word recognition. There

3In an earlier paper (Frauenfelder & Peeters, 1990), we attempted to relate Shillcock’s findings to simulation results obtained with TRACE. We pointed out several problems that complicate any such comparison. First, it is difficult to decide which simulated activation level in TRACE actually constitutes a prediction of activation in an experiment (i.e., are the low activation levels attained by the embedded word sufficient?). Second, it is not obvious which time cycle during the simulation should be selected to determine the simulated activation level for comparison with the experimental RTs to the visual probe. The comparison between simulation and experimental data is further complicated by the fact that the activation patterns obtained in TRACE are determined by the properties of the words and their competitors. These factors were not controlled in Shillcock’s experiments.

are relatively fewer studies of lexical competition; it is more difficult to demonstrate the effects of competition than effects of activation.

**Effects of Lexical Competition**

Some evidence for competition comes from priming studies by Goldinger, Luce, and Pisoni (1989) and Goldinger, Luce, Pisoni, and Marcario (1992); these studies indicated that the recognition of a word is slower when it is preceded by a phonologically similar word than by a dissimilar word. This inhibition is most likely attributable to competition between the prime and target. More direct evidence for competition effects has been provided in a word spotting experiment (McQueen, Norris, & Cutler, 1994). Here, the recognition of a noninitially embedded word (e.g., *mess*) was delayed by its overlapping competitor (e.g., *domestic*). Subjects had to detect these noninitially embedded words in the initial fragment of words (e.g., *domes*) or in control nonword fragments (e.g., *nemes*). The longer spotting latencies for the words embedded in word fragments suggested not only that words with different alignments (carrier and embedded words) are activated simultaneously, but also that the longer carrier word competes with and inhibits the embedded word.

**Effects of Number of Competitors**

Researchers have also studied the effects of competitor set size on word recognition. Marslen-Wilson (1984) showed that subjects' latencies to decide that a speech string was a nonword did not depend on the number of cohort members just before the nonword segment; constant RTs were found independent of cohort size. Conflicting findings were obtained by Jakimik (1979), who obtained evidence for cohort size effects. Slower mispronunciation detection latencies were obtained for nonwords derived from words with a large competitor set (e.g., *complain*) than for those with a small set (e.g., *shampoo*). Unfortunately, the stimuli with large cohorts were mostly prefixed words, a fact that introduces the additional confounding factor of morphological complexity.

A series of experimental studies (Goldinger et al., 1989; P. A. Luce, 1986; P. A. Luce, Pisoni, & Goldinger, 1990) examined the role of the number of competitors with another competitor definition, the *n-count* definition. Here, words with a mismatch in any single phonemic position are competitors. Larger competitor sets produced longer auditory lexical decision and word naming times and also more incorrect perceptual identifications. The inhibitory influence of the competitors is consistent with the predictions made by the authors’ neighborhood activation model (P. A. Luce et al., 1990). To fairly compare these results with the predictions of other models, the effect of cohort and *n-count* competitors must be teased apart more systematically.
A word spotting experiment (Norris et al., 1995) manipulated the number of words consistent with the onset of the second syllable of a bisyllabic nonword. Subjects were required to spot an initially embedded word (e.g., mask in maskukd) that overlapped with these competitors (and that started with sk). The results showed that spotting latencies were slower when there were more overlapping competitors. Using a cross-modal repetition priming procedure, Vroomen and de Gelder (1995) obtained convergent results. Subjects made slower lexical decisions to visually presented targets (e.g., melb) when primed by a spoken sequence (e.g., melkum) with many overlapping competitors than by sequences with few competitors (e.g., melkom).

How do these findings relate to the predictions of different models? The results provide further evidence in support of the TRACE and Shortlist models, which both predict effects of competitor set size, but run counter to the predictions of the Cohort model. There is, however, one additional complicating factor in drawing the negative conclusion about the Cohort model. Cohort II, while excluding competitor inhibition during lexical selection, nonetheless includes certain competitors in the computation of its decision rule. The decision rule establishes the principles for determining when the winning (or perceived) lexical entry is identified. By taking the activation level of the most activated competitor into consideration, this model reintroduces effects of competitors.

**TRACE SIMULATIONS**

We now present a series of simulations that explore the word recognition behavior of TRACE. An important advantage of an implemented model like TRACE over its verbal rivals is its ability to specify the time course of the recognition of individual words. The model produces as its output the activation level across time of the target word and of all the other words in the lexicon. On the basis of these data, the moment of recognition can be specified when some explicit assumptions about the question of how an ultimate choice (with a decision rule) is made. Despite the ability of TRACE to characterize the time course of recognition of every word, we still do not have clear answers to the first two basic questions. Indeed, the complex dynamics created by the interaction of bottom-up and top-down excitation together with lateral inhibition makes it difficult to understand what the competitor set actually is and how it influences word recognition. Although the interactive activation mechanisms underlying these models are quite simple, they can lead to complex and counterintuitive results.

The simulations attempt to characterize this set and to consider how closely it corresponds to the cohort. First, TRACE is submitted to a gating experiment to determine how much bottom-up information is required for word recognition. In the second and third simulations, the competitor set is manipulated to identify which competitors influence target recognition most strongly. Then, we compute the simulated recognition points of a large set of individual words. Several different hypotheses about the competitor set are assessed using these simulation results. By computing and comparing how well the simulation data are explained by predictive measures based on different assumptions about the competitor set, we can understand the basic mechanisms at work in the model. We then examine a more indirect influence of competitors on target recognition via top-down feedback to the phoneme level and the return of this activation to the lexical level.

**Simulation Prerequisites**

In this section, we discuss some prerequisites for modeling word recognition with TRACE and in particular for determining the time course of word recognition it predicts. Our choices about the lexicon and the parameter set are described and motivated. We also consider alternative ways for filling in the missing decision component of the model.

**The Lexicon**

The size and content of the lexicon used in TRACE simulations can be defined as a function of the specific objectives of the simulation experiment. The original TRACE simulations by McClelland and Elman (1986) used a lexicon, Phonolex, that contained 211 lexical entries. These corresponded to all the words in Kucera and Francis (1967) with a frequency of 20 per million or more and containing the 14 phonemes known to TRACE. It is also possible to restrict the lexicon to one or two members to evaluate the influence of a single competitor on the recognition of a target word without the interference of other lexical entries (Frauenfelder & Peeters, 1990). The current objectives are different; we want to study competitor effects in a lexicon that approximates the human lexicon as closely as possible. Therefore, using an American English database of word forms, the largest possible lexicon was generated, in view of the constraints on TRACE’s phoneme inventory. The resultant lexicon, Biglex, contained all 1,024 words that were exclusively made up of the phonemes known to TRACE. In a comparative analysis (Peeters, 1989), this lexicon was shown to resemble the human lexicon more than does Phonolex with respect to average word length and average position of the UP. All the simulations here are performed with (some form of) the Biglex lexicon.

**The Parameter Set**

Connection strengths between units are weighted for each distinct type of connection. These weights are set as values of parameters given to TRACE. The resting level, the minimum and maximum activation values, and the decay rate of the units are also defined in this parameter set. Because the performance
of the model depends on the specific parameter set used, different parameter sets can lead to divergent simulation results. To avoid the criticism that the desired simulation outcome was obtained by parameter manipulation, McClelland and Elman (1986) handpicked a fixed parameter set that they used consistently in all their simulations. This parameter set, claimed by McClelland and Elman to be relatively robust, has been used as a default for studies with TRACE and is well suited to the Phonolex lexicon. When a larger lexicon, Biglex, is used, however, this parameter set was shown not to be appropriate as target words were often not recognized with it (Peeters, 1989). To obtain a better performance, some modifications to this default set were necessary. The new parameter set, listed in Table 4.2, was chosen such that it does not change the basic behavioral patterns of TRACE.

In this new set, the word-to-phoneme excitation parameter is reduced from .030 to .015 and the weight for word-to-word inhibition from .030 to .025. This parameter set is used in all the simulations presented with the Biglex lexicon. The word frequency parameter was turned off for these simulations so that neither word frequency nor competitor frequency played a role.

**The Decision Mechanism**

To determine exactly when a particular phoneme or word is recognized, we need a way of interpreting the activation pattern of the excited units (both target and competitors). McClelland and Elman (1986, p. 21) have proposed a solution for defining phoneme recognition, but no mechanism was offered for word recognition. In order to answer the question posed earlier about the RP of a word in TRACE, we must make two explicit choices. First, we need a decision mechanism to interpret the overall activation pattern produced at each moment in processing and to decide when the recognition of a specific word has taken place. Second, because in TRACE each word type is represented by several tokens (units at different moments in time), we must decide which token(s) of the target word (and its competitors) should participate in the decision process.

Two main procedures can be used to specify the exact moment (in processing cycles) at which a target word is recognized. These differ essentially in which competitors are included in the computation of the RP. We consider two extreme cases here: none or all. In the first case, only the activation level of the target determines the moment of recognition. When its activation level reaches a given threshold level, the target word is considered to be recognized. This procedure represents the most direct way of translating activation into recognition, without having to define secondary procedures. Furthermore, it restricts lexical competition to the activation process. This approach of specifying a threshold value is referred to as the criterion rule.

The alternative competition mechanism is more complex and involves comparing the activation of the target word with that of its competitors. The target word is recognized only when its activation level surpasses that of its competitors by some specified amount. Not the absolute but the relative "goodness of fit" defines recognition. The main problem with this goodness rule procedure is to decide how the relative activation levels of the target and its competitors should be compared and evaluated. The mechanism adopted here is R. D. Luce's choice rule (1959), following McClelland and Rumelhart (1981) in their letter perception model and McClelland and Elman (1980), who used the same mechanism for phoneme recognition in TRACE. It has the form:

\[ R_i = \frac{S_i}{\sum S_j} \text{ where } S_j = e^{k \cdot a_j} \]  

and where \( i \) is the target word, \( R_i \) its response strength, \( j \) a word in the competitor set, and \( a_j \) the activation level of that word. The exponential transformation serves to turn negative values into positive ones, and the multiplication by the constant \( k \) to increase the difference between weakly and strongly activated words. If the response strength (which is a value bounded between 0 and 1) exceeds a certain threshold level, the target word is recognized. Notice that both the response weighting factor \( k \) and the threshold level must be determined in advance because they both influence the exact location of the word's RP.

We exploited both decision rules in determining word recognition durations in the following simulations. The activation threshold level \( c \) for the criterion rule was set to .50. The recognition parameters for the goodness rule were fixed at .9 for the threshold level and 20 for the factor \( k \). With these values, target word recognition is about as fast as possible without allowing nontarget words to be recognized. Different values for \( k \) and \( c \), however, again within reasonable limits, do not alter the simulation results dramatically. Both of these rules produced comparable recognition times in terms of means and standard deviations.
The Target Word and the Competitors

TRACE generates multiple tokens of the each word starting at different moments. In determining a word’s RP, we must decide which of the many tokens of the same word should serve as the target. For a particular target word input, at least one target token is usually sufficiently activated to be recognized (independent of the definition of recognition). Generally this token is the one that is perfectly aligned with the input sequence, because it receives the most excitation from the feature level. An obvious principle for choosing the target unit is to consider only the perfectly aligned tokens. In TRACE, however, both bottom-up and top-down activation also spread to adjacent tokens. As a result, target tokens shifted one time slice to the right often become activated, occasionally even more activated than the perfectly aligned tokens, particularly when the latter are inhibited by the competing words. Since these single time-slice-shifted tokens are clearly not competitors, they were included in the computation of the activation levels of the targets for the goodness rule. As far as the selection of the competitor units included in the computation with the goodness rule is concerned, we considered all nontarget word tokens that overlap with the target to be competitors.

Thus, the simulation results are based on the following choices. For the criterion rule, only the perfectly aligned unit serves as the target. For the goodness rule, both the perfectly aligned and the word token one slice shifted to the right are used as targets. In applying Luce’s choice rule, the (exponentially transformed) response strengths of the target units are added; the resulting sum is then related to the total response strengths of all units, including all overlapping activated competitors.

In the following series of simulations, we address the two questions about the competitor set in TRACE. We begin with two simulations that attempt to characterize the relevant competitor set in TRACE. The remaining three simulations focus on the way in which these competitors influence word recognition.

Simulation 1: Exploring the Members of TRACE’s Competitor Set

As we pointed out previously, the competitor set in TRACE is potentially extremely large. TRACE allows every word to be in contention for recognition at every moment in time. The actual competitor set is much more restricted because of lateral inhibition between words (cf. Frauenfelder & Peeters, 1990). Many word tokens that received bottom-up activation from matching phonemes in the input never become excited above their resting level because they are strongly inhibited by already activated words. It is important to determine which competitors really do become activated during the recognition process. In trying to characterize the competitor set, it is reasonable to ask whether cohort members constitute the critical competitors in TRACE as they do in the Cohort model.

To address this question, we appeal to a simulated version of the gating task (Grosjean, 1980). In this task, subjects are presented successively larger chunks of acoustic-phonetic information and are requested to guess the identity of the partially presented word after each presentation. As already described, several researchers (Grosjean, 1980; Tyler & Wessels, 1983) have used this task in testing human word recognition performance. The RP obtained in these studies corresponds closely to the UP as predicted by the Cohort model (but see Bard, Shilcock, & Altmann, 1988; Grosjean, 1985). If cohort competitors determine the time course of word recognition, we would expect the UP to correspond to the amount of information required by TRACE to recognize the target word. In the following simulation, the TRACE model is subjected to the gating task in conditions roughly comparable to gating experiments with human subjects.

Input. All 977 words in the Biglex lexicon three to eight phonemes in length served as input to the model. A boundary symbol was added at the onset and the offset of each word.

Procedure. Each input word was presented to the model in several separate simulation runs. On the first run, only the initial phonemic segment of the target word (preceded by the boundary symbol) was given as input; on the second run, the first two segments were presented, and so forth through the final word boundary symbol. Both the activation levels and the response strength levels for every input were recorded for each run, and the recognition points were computed according to both the criterion rule and the goodness rule. This procedure was used to establish the minimal amount of information (measured in number of segments) required for recognition. The position of the critical phoneme that allowed recognition is referred to as the Isolation Point (IP). One special class of words, those that do not have a UP before their offset, was analyzed separately. Biglex served as the lexicon.

Results. To get an initial impression of the time course of word recognition, we selected a subset of items that consisted in all 21 words seven phonemes long with their UP in the fourth position. Figure 4.2 shows their average activation (ACT) curves, and Figure 4.3 shows the response strength (RS) curves as a function of the number of phonemes present in the input (ranging from one to seven including the final word boundary symbol).

The figures show that an input from the word onset up to and including the UP (i.e., the first four phonemes) is both necessary and sufficient for recognition according to the predefined thresholds for the two rules (c = .50 in Figure 4.2, and q = .90 in Figure 4.3). Figure 4.2 shows that the target does not become
excited until well after the UP. Moreover, recognition does not take place immediately at the UP. Both with and without additional input after the UP (cycle 30), processing continues after this point for many (20 to 50) cycles until the word is finally recognized. More specifically, the RP for the full input is located at cycle 55 in both figures, whereas, for the input of four phonemes, it is located much later at cycles 78 and 81, respectively.

The relation between the UP and the IP for all the words in this simulation is shown in Table 4.3. This table gives the number of words that have a specific temporal relation (before, at, and after) between the location of their IP and their UP. The words are classified together on the basis of their length and their UP positions.

For each length–UP class, we listed from left to right the number of words whose IP occurs before the UP, precisely at the UP (bold), or after the UP. In most cases (almost 85% for the criterion rule and over 89% for the goodness rule), the IP corresponds exactly to the UP. For about 7% of the words presented, one or two additional phonemes were required for recognition (with both decision rules). Rather surprisingly, in some of the words (8% and 4% for the two rules, respectively) less information than that defined by the UP phoneme was sufficient for recognition. This was often the case for words with a UP after their offset, that is, those words that are only unique when the word boundary symbol is presented (shown in the last column of the table). The table reveals the special nature of this class of words; almost all are recognized on the basis of partial sensory input, with the final one to three phonemes of input not required.

**Discussion.** Simulation 1 shows that TRACE recognizes words on the basis of partial information. More precisely, the amount of input both necessary and sufficient for recognition is best defined by the location of the UP. This finding tentatively answers our first question about the nature of the competitor set. Indeed, it suggests that this set does not include noncohort competitors, that is, those words that match the target but are misaligned with it or those words that are aligned but that mismatch minimally. These lexical competitors appear not to be decisive in determining the amount of information required for recognition. If these words actually exerted a major influence, the UP effects would be much weaker. Although more simulations are certainly needed, the present results suggest that the competitor set in the TRACE model at a particular moment can be closely approximated by the set of all initially aligned words that match the input sequence, that is, by cohort competitors.

In Cohort I, the sensory information from word onset through and including the UP phoneme is by definition sufficient for recognition. This relation between the UP and word recognition was also observed in TRACE for more than 80% of the words with their UP before their offset. The models appear to differ in their predictions about the remaining words, for example those with their UPS after offset. Although the Cohort model has never been specific about the processing of these words, we can assume that they are recognized only after the entire word and the following segment or silence.
has been processed. In contrast, the TRACE model shows that most of these words are recognized on the basis of less information than that delimited by the (completed) UP phoneme. This early recognition is attributable to the way in which inhibition is implemented in the model; the amount of inhibition between two words is a function of the number of phonemes in which they overlap. Consequently, in TRACE, shorter words are less inhibited by their competitors than are longer words (McClelland & Elman, 1986). Because words with the UP after their offset are generally the shortest in their competitor set and are least affected by inhibition, they are more likely to reach the activation level required for recognition. This simulation result suggests that such short initially embedded words are likely to be recognized spuriously (Dennis Norris, personal communication), a prediction not confirmed to our knowledge empirically, for instance, in slips of the ear corpora (collections of slips of the ear data).

The models also appear to differ in their predictions about recognition latencies for all the words. Unlike the Cohort model, TRACE does not recognize the target words immediately at the UP, but only much later toward the end of the word. Moreover, the recognition latencies for the same word differ as a function of how much sensory input is supplied. The simulation results show that more than 25 additional cycles are required to recognize a word when only the initial part including the UP phoneme is presented as opposed to the entire word. The Cohort model does not predict any differences as a function of the amount of information presented after the UP.

### Simulation 2: Comparing Cohort and Noncohort Competitors

A more direct way of establishing which words act as competitors in TRACE is to manipulate the lexica used in the simulations. It is straightforward to eliminate specific types of competitors from the lexicon and to compute the recognition latencies for a set of target words with these reduced lexica. In the following simulation, we compare the recognition performance of TRACE with cohort and noncohort competitors omitted from the lexicon. This comparison includes four main conditions: baseline with the full Biglex lexicon, lexica without final cohort competitors, lexica without mismatching competitors, and lexica without misaligned competitors. By subtracting the baseline activation curves from each of the reduced conditions, we can evaluate the influence of specific competitor types. The larger the differences with respect to the baseline, the more influential the particular competitor set is on word recognition.

**Input.** The same 21 input words as in the preceding simulation (seven-phoneme words with the UP at the fourth position) served as input.
**Procedure.** The activation curves for each word were obtained for the six different lexica. These lexica were all based on Biglex with specific competitors eliminated. The six lexica are illustrated using one of the target words, *surplus* (*/sʌrplʌs/ (UP = p)). The five reduced lexica can be assigned to one of three classes: final cohort competitors, mismatching competitors, and misaligned competitors. These lexica are described here along with the average number of entries they contain.

Lexicon 0 (Baseline): Biglex lexicon (1024 entries).
Lexicon 1 (Final cohort competitors): Biglex lexicon excluding all words (e.g., */sʌ₃r.../) that are cohort members at the phoneme before the UP; 1,019 entries on average.
Lexicon 2 (Mismatching competitors 1): Biglex lexicon excluding all words that match the target in the first two phonemes but mismatch from the third phoneme position onward (e.g., */sʌ.../); 1,066 entries on average.
Lexicon 3 (Mismatching competitors 2): Biglex lexicon excluding all words that match in first and third phoneme positions but that mismatch in second phoneme (e.g., */s_.../); 1,017 entries on average.
Lexicon 4 (Mismatching competitors 3): Biglex lexicon excluding all words that match in second and third phoneme position but that mismatch in the first phoneme (e.g., */_s.../); 1,002 entries on average.
Lexicon 5 (Misaligned competitors): Biglex lexicon excluding all words that are misaligned to the right with the target. The first and second phonemes of these competitors match the second and third phonemes of the target (e.g., */_s.../); 1,016 entries on average.

**Results.** The activation curves for all 21 input words were averaged for each lexicon condition. Then the curves from the conditions with the reduced lexicon were subtracted from the baseline results to produce the five differential activation curves shown in Figure 4.4.

This figure shows that Lexicon 1 (final cohort competitors excluded) produced the largest differential activation with respect to the baseline lexicon. The differential activation begins to increase around the UP and reaches its maximum value toward the end of the word. Lexicon 2 (the first mismatch condition) produced some differential activation that reaches its maximum earlier, around the UP. The remaining lexica, which represent mismatching or misaligned conditions, did not differ from the baseline.

**Discussion.** This simulation presents a clear picture of the influence of different competitors on the simulated word recognition behavior of TRACE. The results showed that final cohort competitors had the largest impact on the recognition of targets. The only other competitors that produce an effect were those eliminated in Lexicon 2, matching in the first two phonemes but mismatching in the third. These words are in fact nonfinal cohort members that drop out of the cohort one segment earlier than did the final cohort members. They lead to earlier differential activation than did the final cohort members (Lexicon 1) because of their greater number. Taken together, the results of Simulation 2 are entirely consistent with the idea that cohort competitors play a decisive role in determining TRACE's predicted time course of word recognition.

A caveat should be added about the influence of nonaligned targets. In the present simulation with Lexicon 5, we included only the competitors that were misaligned to the right of the target. Thus, none of the competitors began before the target as would be the case if the targets were embedded in a longer word sequence. In this case (see Frauenfelder & Peeters, 1990), the matching left-aligned competitors would completely dominate the target and prevent it from ever being activated. The situation of competitors aligned to the right and to the left is completely different. We are concerned here primarily with the former.

TRACE behaves like Cohort I in terms of its core competitor set—surprising in view of the differences in their underlying activation mechanisms and in their apparent tolerance of mismatch. For TRACE, the fact that two similar phonemes can both be activated, albeit to different degrees, by a given input because of their shared distinctive features and the fact that these phonemes can in turn activate the words that contain them suggests that close rhyming
words should be activated and become competitors. This result did not emerge from the simulations because the matching word became much more activated than did its closest matching competitors and inhibited these words. Thus, on the basis of these first two simulations, we can conclude that only matching and aligned competitors become strongly activated in TRACE.

In what follows, we shift our attention to the question about the influence of the competitors on target recognition. We may expect differences between the Cohort and TRACE models to emerge more clearly.

Simulation 3: Exploring Other Properties of the Competitor Set

This simulation seeks to answer the question about how the competitor set influences the recognition of a target word. As in the previous simulation, the lexicon or competitor set is manipulated while the input is held constant. Instead of evaluating the influence of the different competitors, we now focus on one type, the aligned and longest matching competitors, that is, the final cohort members. More specifically, we examine how the size and properties of these cohort members influence the recognition latencies.

**Input.** The input to the model is the seven-phoneme word *surplus*, phonetically /ˈʃɜːrpəl/; The competitor set of this target contains exactly 15 members in Biglex, including *sir*, *search*, *circle*, *surcease*.

**Procedure.** A series of simulation runs was conducted for the same input target word but with different lexica (all based on Biglex). The first lexicon excluded all 15 final cohort members (like Lexicon 1 in the previous simulation). For the second run, one cohort competitor was reinserted into the lexicon. With every successive presentation of the target word, another member of the competitor set was added to the lexicon in random order. The procedure was repeated 15 times until the lexicon was complete again. For every run, the activation curves of the target word were recorded. In the final run, the full Biglex was used as a baseline with which the other results could be compared.

**Results.** Sixteen activation curves were generated by this simulation. On the basis of these curves, we computed the RS with the procedure previously described. Differential RS curves were then produced by subtracting the curves of the test lexic from that of the baseline lexicon, Biglex. The results are plotted in Figure 4.5.

An increase in the number of competitors leads to a reduction in the activation level (and accordingly the RS level) of the target word. Every additional competitor has some inhibitory effect on the target word's activation level and, consequently, on its RP. Hence, we can conclude that the competitor set size is an important factor in determining a word's RP.

Figure 4.5 also reveals another striking phenomenon. Some competitors appear to exert a larger inhibitory effect than do others. In particular, the seventh competitor shows a disproportionately large inhibitory influence on target recognition. This competitor is the word *sir,* which, unlike the others, is completely embedded in the target word. The competitors that are embedded initially in the target thus have a special status in the competitor set.

**Discussion.** This simulation has identified two of the competitor set's properties that influence the activation level of the target word: the number of cohort members and the presence of embedded words. The results show that each additional cohort member contributes its own inhibitory influence on the target, but the inhibitory influence of competitors is not equal. Embedded competitors—as in Simulation 1—become slightly more activated than do other competitors. The amount of inhibition a word receives is a function of its own length. Shorter embedded words get less inhibition from other words in the lexicon than do longer words, because they overlap with their competitors in fewer phonemes. As a result, these competitors become more activated and can exert a stronger inhibitory influence on target recognition.

This simulation has revealed two differences between the TRACE and Cohort models. First, an effect of cohort size in TRACE results from the lateral inhibition between cohort members. Second, an effect of embedded words is clearly linked to specific assumptions about how the amount of
inhibition between words varies as a function of their length and overlap. Because this simulation was based only on a single word, it might be premature to draw general conclusions about the influence of the competitor set on target recognition. The next simulation was designed to overcome this limitation and to examine these properties of the competitor set more systematically with a larger set of words.

**Simulation 4: Quantitatively Relating the RP to the Competitor Set**

The results of the previous simulations can be summarized as follows. Simulations 1 and 2 showed that those words that are aligned with the input and that match it, the cohort members, best define the competitor set in TRACE. Consequently, the RP of a target word input is highly dependent on its UP. Moreover, Simulation 3 indicated that the competitors' influence varied as a function of their number and their embedded status. The findings are next explored in detail in a quantitative analysis. The RPs of a large number of target words are determined in a simulation, and these simulated recognition data are then compared with predictive measures based on different assumptions about the competitor set.

**Input.** The input to the model was same set of 977 words from Biglex as those used in Simulation 1 (three to eight phonemes in length). Each complete word, bounded on both sides by the word boundary symbol, was presented to the model.

**Procedure.** The RPs were determined for every word according to both the criterion rule and the goodness rule procedures already described. In the preceding simulations, the individual recognition latencies obtained with these two rules differed slightly (with greater variance for the criterion rule data), but essentially no difference in the overall patterns was obtained with the two rules. We therefore discuss only the following results for the goodness rule, but refer to the results for the criterion rule wherever interesting differences arise.

**Results.** The words with the same length and UP position were clustered into groups. For each group, the average RP was calculated with the goodness rule. Figure 4.6 shows the average RP of each group of words with the same UP position as a function of word length. All data points in one UP class are connected by lines. Words with their UPs after their offsets are not included in this figure.

---

4. **SIMULATING THE TIME COURSE OF SPOKEN WORD RECOGNITION**

![Graph](image)

**FIG. 4.6.** Average RPs, in cycles for words with different UP positions as a function of their phonemic length.

The figure shows a strikingly regular relation between the positions of the UP and the RP for each length class. Without exception, words with earlier UPs are recognized faster, a finding that provides additional evidence in support of our previous conclusion that the cohort is the competitor set in TRACE. This figure also shows that word length has a clear influence on recognition latencies. For most UP positions, there was a U-shaped curve for the RPs across the different word lengths examined. To eliminate the effects of word length, we restricted ourselves in the following analyses to the longest (seven- or eight-phoneme) words in the lexicon. This group of 114 words represented a large range of UP positions.

We computed the RPs, average UP position, and word offset for each UP class. The latter two were defined as the cycle at which the amount of activation fed by the corresponding distinctive feature units into the phoneme unit representing the UP or into the word offset boundary symbol reached its respective maximum value. The average RP (at 58.61 cycles) comes just after the word offset (after 57 cycles), but 22 processing cycles after the UP (at 36 cycles). If we convert processing cycles into real time (25 ms per cycle) following McClelland and Elman (1986, p. 14), we arrive at a delay of 550 milliseconds from the UP to the RP. Nonetheless, a high correlation ($r = .78$) was obtained between the position of the UP and the position of the RP. In addition, 61% of the variance in RPs can be explained by the UP. In what follows, we examine the role of UP more closely.

---

$^4$Interpolation techniques, using the Lagrange polynomial (for details about this procedure, see Daniels, 1978), made it possible to determine the RP with considerable precision.

$^5$This value can be derived from the position of either the unique or word-boundary phoneme by adding one for the initial word boundary phoneme and multiplying this outcome by six, which is the length of the phoneme in cycles.
A Quantitative Evaluation of the Properties of the Competitor Set

In this section, the results of Simulation 4 are analyzed in greater detail using measures that predict the RP of the target word on the basis of a different assumption about its competitor set. The measures are based on a lexical-statistical analysis of three different properties of the target words: the UP position, the number of cohort members, and the presence or absence of embedded cohort members. By comparing the ability of these measures to predict the location of the simulated RPs, we hope to gain further insight into the nature of the competitor set and the interaction between competitors and targets in TRACE.

To predict the observed RPs, we looked for the best fitting values for diverse parameters with the help of the parameter search procedure STEPIT (Chandler, 1969; Massaro, 1987). The program finds an optimal set of parameters on the basis of the observed RPs, such that the root mean square deviation (RMSD) between the predicted and observed values is at a minimum. To be consistent and to use the same analysis technique on all predictor models, STEPIT was applied to the (linear and nonlinear) models tested.

Analysis 1: The Uniqueness Point as a Predictor

The first predictor to be examined is based on the UP. The predicted RP (P(RP)) for a particular word w is expressed in Equation 2. This equation includes two main terms.

\[ P(RP_w) = B + N_w \cdot I \]  

(2)

First, it uses a measure of basic processing time \((B)\), which can be viewed as the recognition duration of words without competitors. Second, there is a measure of competitor cost for the UP predictor, the product of the number of phoneme positions for which the target word has competitors \((N_w \text{ or UP } - 1)\) and a parameter \((I)\) representing the increase in the recognition duration for each position that there are competitors. The two parameters, \(B\) and \(I\), are free, and \(N_w\) is an empirical value.

The parameter search procedure produced the following values for \(B\) and \(I\) respectively as shown in Table 4.4: 45.23 and 3.24. The obtained RMSD of 2.79 is fairly low. The relation between the predicted and simulated RPs is shown in Figure 4.7.

The use of a constant \((I)\) to represent the increase in recognition duration per position seriously limits the predictive power of this measure. Therefore,

\[ P(RP_w) = B + \sum_{i=1}^{N_w} I_i \]

(3)

we used an alternative equation that varies \((I_i)\) over positions \(i\), ranging from the first position to the position immediately before the UP:

The number of free parameters in this measure is larger than in Equation 2; the exact number depends on the number of positions before the UP in the target words. For the present stimulus set there are seven free parameters (one \(B\) and six \(I_i\)). The new RMSD measure of 2.67 is considerably lower.
measure based on the competitor set size, we used the log transform of cohort size. Further, we included a weighting factor that represents the influence of the cohort size at each sequential position. We expect that the size of the cohort in later positions played a larger role in determining the RP; large cohorts in later positions should lead to the longest delays in recognition.

To evaluate the role of competitor set size, we used an Equation 4, which differed from Equation 3 in only one term, \( S_w \), representing the (log) competitor set size of word \( w \) on phonemic position \( i \). This parameter serves to weight \( I_i \):

\[
P(RP_w) = B + \sum_{i=1}^{N} I_i \cdot S_w
\]

(4)

Because \( S_w \) can be derived from an analysis of the lexicon, the number of free parameters remains the same as in Equation 3. The values of the parameters generated (cf. Table 4,4, col. 3) are useful for interpreting the results. The free parameter, \( I_i \) for the first two word-initial positions (both 0) is easy to interpret; it indicates that the initial cohort size does not contribute to the prediction of recognition latencies. The contribution of this parameter increases for later positions and reaches its maximum just before the UP. The evolution of cohort size shows just the opposite tendency; in early positions it is generally extremely high and decreases sharply in later positions (as for a larger computerized lexicon; cf. Marslen-Wilson, 1984). This analysis suggests that the number of cohort competitors is an important factor for later positions when only relatively few competitors are still in contention.

The relation between the predicted and simulated RPs is shown in Figure 4.8.

It is interesting to compare the predictive measure based on competitor set size with that based on UP. The former led to a better fit; the RMSD dropped to 2.49. The improved fit of the predictive measure that includes cohort size confirms that this factor plays a role in word recognition in TRACE. This result is clearly different from that of the Cohort model that specifically denies the relevance of this factor. The impact of embedded words constitutes a second difference with the Cohort model. In the following analysis this factor is investigated.

Analysis 3: The Embedded Status of Competitors as a Predictor

This analysis investigates the impact on word recognition of competitors that are entirely embedded in the target word. Every competitor was classified in terms of whether it was completely contained in the target word.
There were two different categories of competitors: one with completely embedded competitors and one with words that mismatched the target words in some later position. We then determined the latest positions at which the competitor set still contained a representative of either category. In Equation 5, these two positions are given by \( P_u \) for embedded and \( Q_u \) for nonembedded competitors:

\[
P(RP_u) = B + \sum_{i=1}^{P_u} I_i \cdot E + \sum_{i=1}^{Q_u} I_i \cdot (1 - E)
\]

The variable \( E \) represents the weight given to embedded words and \((1 - E)\) the weight for nonembedded words. The variable \( E \) is an additional parameter in the model, so that the total number of parameters (eight) increases by one in comparison with Equation 3.

This measure predicts the simulated recognition latencies quite well. The RMSD is reduced to 2.23, better than the value obtained with the cohort size predictor. The value of \( E \) (expressing the contribution of embedded words) is quite high (\( E = .654 \)), almost twice as high as that for nonembedded words. This result confirms that the embedded competitors have the biggest impact on the time course of word recognition.

We can derive another equation (Equation 6) to include both embedding and the cohort size:

\[
P(RP_u) = B + \sum_{i=1}^{P_u} I_i \cdot S_{wd} \cdot E + \sum_{i=1}^{Q_u} I_i \cdot S_{wd} \cdot (1 - E)
\]

In comparison with Equation 5, no free parameters are added. The gain of this predictive measure over the previous one is minor; there is only a small reduction in RMSD (from 2.23 to 2.20). The parameter \( E \) also increased slightly to .676 and results in an even higher weight for embedded words.

**Summary: The Role of Competitor Set Size and Embedding Status**

In the preceding, we evaluated the contribution of different properties of the competitor set to the simulated recognition behavior of TRACE by comparing the quantitative fits obtained with different predictive measures based on these properties. This analysis showed a good fit between the predicted and the obtained latencies when the position of UP was used. Even better fits were obtained with two other factors, competitor set size and embedding status. Although these measures are able to predict the recognition latencies of words with considerable accuracy, some unexplained variance still remains. Certainly some of this residue is attributable to the way in which the front end is implemented in the model, in particular the values that determine how many phonemes are activated by specific features in the input.

One additional factor that could also contribute to the unexplained variance is the top-down feedback from the lexical to the phoneme level and the return of this activation to the lexical level, the so-called resonance effect. Comparisons of the fits for the recognition latencies that we obtained in simulations with and without top-down lexical to phoneme feedback suggest that this factor is relevant. These results systematically showed weaker correlations when top-down was included than when it was excluded. In the following simulation, this "resonance" effect of top-down feedback on the lexical level is examined in greater depth.

**Simulation 5: The Effect of Top-Down Excitation on Word Recognition**

To arrive at a complete characterization of the influence of competitors on word recognition, we must also investigate the way in which competitors affect target word recognition indirectly via top-down lexical feedback. Unfortunately, little attention has been paid to this problem. Indeed, only the effects of top-down feedback on phoneme recognition have been investigated with TRACE. Top-down feedback has generally been assumed to facilitate word recognition via two related activation pathways. First, top-
down excited phoneme units making up the target return their higher activation to this lexical target. Second and more relevant for the present discussion, the activated lexical competitors excite the phonemes they share with the target. Again, this increase in the activation levels of the phonemes contained in the target is passed back to the lexical target unit. This so-called gang effect, first identified by Rumelhart and McClelland (1982) in their interactive activation letter perception model, should lead to faster recognition latencies for target words in TRACE. The amount of facilitation produced by the gang effect should increase as a function of the competitor set size; the more competitors there are, the larger the effect.

If indeed these facilitatory between-level effects accelerate word recognition, they would work against the inhibitory in-level competitor effects that slow down target recognition through lateral inhibition. The conflict between competitor and gang effects might well account for some of the unexplained variance previously alluded to. In the following, we compare the simulated recognition latencies of a set of words (matched in length and position of UP) with and without top-down feedback. We expect shorter word recognition latencies in simulations including top-down feedback because of the activation pathways already described. Furthermore, the top-down effect is presumably more beneficial for words with large cohorts because of gang effects.

**Input.** The input to the model is the same 21 words (seven phonemes; UP in fourth position) as used in Simulations 1 and 3. We also distinguished the input words as a function of the number of competitors they had before the UP.

**Procedure.** The RPs were computed for all the input words with the top-down excitation parameter turned off (reset to 0). The results were then compared with the word RPs using the default parameter set (with feedback).

**Results.** The results of this simulation are plotted in Figure 4.9. Every input word is represented as a function of its RP without the top-down feedback on the horizontal axis and with the top-down feedback on the vertical axis.

The figure shows no consistent pattern of top-down facilitation on word recognition. Faster recognition latencies for words with top-down facilitation occurred for only half the words; for the other half, word recognition was faster without top-down feedback in the model. The number of competitors does not seem to have a consistent effect; words with many competitors fall into both categories.

**Discussion.** The results of this simulation did not confirm the predicted facilitation of word recognition by top-down lexical to phoneme feedback. We expected all the words in the simulation to be recognized faster with top-down feedback than without. In addition, the greatest decrease in recognition latencies was predicted for targets with many competitors, because of the facilitatory gang effect. Our comparison of recognition latencies with and without this feedback revealed that only some words benefited from top-down feedback; others were either unaffected or were slowed down. Moreover, the prediction of greater facilitation for words with large competitor sets was also not confirmed. The two words with the most competitors were the most adversely affected by the top-down feedback.

In the introduction to this simulation, we presented two facilitating pathways by means of which word recognition could be accelerated by top-down feedback. To explain this mixed pattern of results, we must identify alternative pathways that slow target recognition. There are at least two such pathways. According to the first, excited competitors gain in activation by a resonance effect via the phonemes they contain. Second, a target word also activates the shared phonemes contained in their competitors; these phonemes in turn increase the activation of the competitors. In both cases, the more excited competitors can inhibit the target and slow its recognition.

These two inhibiting pathways compete with the two facilitating pathways presented earlier. To understand the complex flow of activation via the diverse connection pathways, we examined other properties of the competitor set to help explain the complicated pattern of results. We were unable to predict whether the top-down feedback would lead to faster or slower
word recognition on the basis of a detailed analysis of the competitor set and of the recognition of individual words. This fact illustrates the complexity of the activation flow in TRACE and the difficulty of understanding the effect of the mechanisms at play in the model.

CONCLUSIONS

In this chapter, we examined several leading models of spoken word recognition, including both a verbal model, Cohort, and two computational models, TRACE and Shortlist. We have compared the answers they provided to the questions posed about lexical activation and selection in light of findings reported in the experimental literature. A series of five TRACE simulations was presented to shed light on the consequences of the specific interactive activation mechanisms, lateral inhibition and top-down facilitation, that this model assumes. In particular, we have attempted to identify the competitor set in TRACE and to determine how different properties of this set determine the time course of word recognition.

These five simulations have shown that the three questions about word recognition can be answered in a direct and straightforward manner for individual words when the necessary assumptions are made about the content of the lexicon and about the decision rule that interprets the activation of the target word (and its competitors). It was a major challenge to generalize from the computer model's performance on specific lexical items to its underlying behavior and then to predict its behavior on novel items.³

We briefly summarize the answers provided by the present simulations to these questions and consider the relative advantages and limitations of verbal and computer-implemented models. To address the question about the members of the functional competitor set, we distinguished two main classes of competitors: cohort and noncohort competitors. The cohort competitors—assumed to be involved in word recognition by the original Cohort model—refer to the set of words that are aligned with the target and match it completely at the moment being considered. Noncohort competitors refer to partially matching or partially aligned words. Taken together, the first two simulations indicate that for TRACE—as for Cohort I—the cohort competitors constitute the core of the lexically activated set.

This conclusion has an obvious but nonetheless important corollary. The fact that the initially mismatching competitors are not very activated in TRACE means that an initially mismatching input does not activate a target word sufficiently for recognition. In other TRACE simulations (Goldman, Frauenfelder, & Content, 1997), this question was explored directly by manipulating several factors: the amount of initial phonological mismatch (none, single, none distinctive feature), and large (at least three features), the lexicon size (Phonolex: 245 words vs. Biglex: 1,045 words), and the parameter settings (original vs. modified top-down and lateral connection strengths). Simulation with the original lexicon (Phonolex) and the original parameter set confirmed that TRACE does not systematically recognize words with onset mismatches. Indeed, the recognition rates observed for small and large mismatches was 75% and 57%, respectively, in contrast to the 100% recognition rate for the matching inputs. Even more striking was the fact that for the larger lexicon, Biglex, the recognition performance of the model plummeted to below 25% even in the case of the close mismatches. In other words, the original and central claim that TRACE recognizes words despite small initial phonological mismatches (that is, sbigarette is recognized as cigarette) appears to be incorrect. This claim that TRACE tolerates minor mismatches was reported by McClelland and Elman (1986), apparently on the basis of a simulation with a specific item (bleeasant > pleasant). In contrast, these simulations also show the inadequacy of item-based conclusions and the necessity for more systematic simulations. On the basis of our simulations, we can conclude that TRACE does not recover from initial mismatches easily or efficiently.

This result is not in agreement with the experimental findings previously reported. These findings suggest that the human listener is relatively flexible when confronted with mispronounced input and can recover the intended word despite the presence of minor initial mismatches. We expect that it should be possible, nonetheless, to simulate such behavior with TRACE using other parameter settings. As we further suggest, these adjustments might well include eliminating top-down influences. Such adjustments of parameters in interactive models of this type raise some important methodological issues. How and when can parameters be modified in a principled modeling approach?

The second question addressed in this paper concerns the specific properties of the cohort competitor set that determine the time course of word recognition. Simulations 3 and 4 (Analysis 2) revealed that the cohort set size influences the location of a word's RP. This prediction received some confirmation from experimental studies that showed a similar relation between word-spotting latencies and the size of the competitor set (Norris et al., 1995). A second property of the competitor set was also shown to determine the simulated time course of word recognition, namely, the presence of embedded cohort members that are onset aligned. Simulation 4 (Analysis 3) demonstrated that such embedded words have a stronger negative impact on the recognition of target words than do noneembedded cohort members. We know of no experimental studies showing that word recognition is slowed by the presence of such embedded words as TRACE predicts. It would be important, therefore, to test this prediction empirically.

³In some sense, computational and verbal models are complementary; for the latter, general predictions are easier to generate than are specific item-based ones.
In Simulation 4, we attempted to evaluate quantitatively the contribution of these properties of the competitor set (including cohort size and presence of embedded words). We devised measures based on a statistical analysis of the lexicon used by TRACE. This research illustrates that lexical statistical analyses of TRACE's lexicon can provide a useful tool for understanding the complex behavior of this model. These measures predicted the recognition latencies with considerable accuracy. In fact, one such measure could account for about 76% of the variance in the RTs. The fit was not perfect; there were several residual effects to be explained despite the relatively large number of free parameters used.

Simulation 5 showed that resonance via top-down lexical feedback had an inconsistent effect on word recognition such that word recognition is not always accelerated but sometimes delayed by this effect. The effect of such resonance on lexical processing has attracted little attention because it has generally been assumed that improved phoneme recognition caused by top-down feedback automatically translates into faster lexical recognition. The present results suggest that this mechanism does not significantly contribute to the performance of TRACE as a word recognition model, and thereby raise some doubts about the fundamental utility of the top-down feedback. In fact, additional simulation results (Goldman et al., 1998) have shown that TRACE deals better with initial mismatches when top-down feedback is turned off entirely. Because it is precisely when there is some distortion in the signal that top-down information is thought to be most urgently needed, this result poses a real problem for the model.

To conclude, we would like to turn our attention to a deeper issue about the use of computer-implemented models in studying lexical processing. Our understanding of word recognition has progressed considerably in the last 2 decades. Psycholinguistic research in this area has identified many factors that play a role in lexical processing. To include these different factors and their complex interaction in a theoretical model, researchers have been forced to construct models of increasing complexity. Thus, the limits of verbal models have now apparently been reached, and computer technology is required to consolidate and integrate the acquired knowledge.

One of the main advantages of computational models—in contrast to a verbal model like the Cohort II—is their ability to make precise quantitative predictions about the time course of spoken word recognition. Although TRACE can generate activation curves and word recognition latencies for each and every word in its lexicon, it is still difficult for a modeller using TRACE to predict the simulation outcome for a specific input. In fact, TRACE occasionally produces some counterintuitive results, as we have seen for the effect of initial mismatches and for that of top-down feedback. This state of affairs presumably will become even worse as the number of factors included in the model increases. In the present simulations, we purposely did not include the variable of word frequency that—although important in determining the time course of word recognition—would have made the results even more difficult to interpret.

Even in its simplest form, TRACE retains some characteristics of the black box that it was intended to explain. Part of the difficulty lies in understanding the complex interaction between the relatively simple processing mechanisms (bottom-up activation, lateral inhibition, and top-down activation) underlying interactive activation theory. Furthermore, it is a challenge to establish the extent to which the obtained results are attributable to theoretically motivated assumptions about the model's mechanisms or to mere implementation decisions. Indeed, in implementing these mechanisms, a modeller is forced to make somewhat arbitrary choices (e.g., about the definition of features, the use of a phoneme level, the parameter settings) that can all affect the output of the model. Clearly, it is important to determine how much the model's observed behavior is really a true reflection of the underlying theory being proposed.

Our research strategy has been to treat TRACE essentially as a human subject—or even an animal system as McCloskey (1991) suggested—and to examine its behavior in different experimental situations. In doing so, we have manipulated the parameter sets, the nature of the lexicon (moving from a highly restricted lexicon to the largest lexicon that TRACE can handle), and the properties of the input. The advantage of such tightly controlled, albeit artificial, simulations is that they make it possible to reduce the complexity of the model's behavior. In this fashion, we have collected behavioral data from the model to get a better grip on what TRACE is actually doing. Our attempts to predict the individual recognition latencies using properties of the competitor set have helped us to identify important factors. Using a predictive measure based on the nature and number of competitors, we have obtained reasonable fits between the observed and the predicted results (predicting 76% of the variance). More work is clearly required to link these findings to the underlying mechanisms of the model and to identify the contribution of unmotivated implementation choices. We conclude by pointing out the clear danger in this approach of being sidetracked and of focusing too closely on a computational model's behavior. This effort is worthwhile only when the real objective, that of understanding human listeners, remains central. Nonetheless, identifying the behavior consequences of the different assumed basic mechanisms constitutes an important step in computer modeling, a step that requires systematic and detailed simulations like those presented here.

ACKNOWLEDGMENTS

Thanks go to Alain Content, Jan-Peter de Ruiter, Ton Dijkstra, Jean-Phillipe Goldman, Jonathan Grainger, and Dennis Norris for their comments on an earlier version of this chapter. Support for this research was provided by
grants from the Swiss National Fund for Scientific Research (Grant 11-39553.93 and Grant 1113-049698.96).

REFERENCES


---

**CHAPTER FIVE**

**MRM-p: An Interactive Activation, Multiple Readout Model of Orthographic and Phonological Processes in Visual Word Recognition**

Arthur M. Jacobs  
*Philipps University of Marburg*

Arnaud Rey  
*Laboratoire de Neurosciences Cognitives - CNRS*

Johannes C. Ziegler  
Jonathan Grainger  
*CREPO-CNRS & Université de Provence*

*The world is worded before it is sentenced.*  
— Variation on a theme by Humboldt

**READING, WORD RECOGNITION, AND THE LEXICAL DECISION TASK**

*The world of words is just as wondrous as the world of syntax, or even more so. For not only are people as infinitely creative with words as they are with phrases and sentences, but memorizing individual words demands its own special virtuosity.*  
— S. Pinker, *The Language Instinct*

The subject of this chapter is reading, one of the finest achievements of human civilization and one of the most complex activities of the human mind. Explaining the why's and hows of reading skill represents an outstanding intellectual challenge for cognitive scientists. Word recognition is the fundamental process underlying reading skill; it provides a favorable focus for experimental reading research. At the level of word representations, all lower and higher level processes involved in reading seem to meet. Word repre-