

Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition

Delphine Dahan

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

James S. Magnuson, Michael K. Tanenhaus,
and Ellen M. Hogan

University of Rochester, NY, USA

Participants' eye movements were monitored as they followed spoken instructions to click on a pictured object with a computer mouse (e.g., "click on the net"). Participants were slower to fixate the target picture when the onset of the target word came from a competitor word (e.g., *ne(ck)t*) than from a nonword (e.g., *ne(p)t*), as predicted by models of spoken-word recognition that incorporate lexical competition. This was found whether the picture of the competitor word (e.g., the picture of a neck) was present on the display or not. Simulations with the TRACE model captured the major trends of fixations to the target and its competitor over time. We argue that eye movements provide a fine-grained measure of lexical activation over time, and thus reveal effects of lexical competition that are masked by response measures such as lexical decisions.

It is now generally accepted that as listeners attend to a spoken word, they simultaneously entertain multiple lexical candidates, which compete for recognition (see Frauenfelder & Floccia, 1998, for a review). However, the mechanism by which competition among active candidates is realised and resolved remains controversial. In some localist connectionist models, such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994), word candidates compete with each other via inhibitory lateral connections. Thus, the activation of a lexical candidate at a given point in time in the

Requests for reprints should be addressed to Delphine Dahan, Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands.

E-mail: delphine.dahan@mpi.nl

This work was supported by an NSF grant SBR-9729095 to M.K. Tanenhaus and R.N. Aslin, and by an NSF GRF to J.S. Magnuson.

recognition process is determined both by its fit with the input and by the activation of other candidates. In contrast, the Cohort model (Marslen-Wilson, 1987, 1990) does not assume lateral inhibition. The activation of a lexical candidate is affected only by its goodness of fit with the input and not by the activation levels of competitors; lexical competition takes place only at the decision stage of recognition. More recently, Gaskell and Marslen-Wilson (1997, 1999) have proposed a distributed model in which word recognition is viewed as the activation of a set of features that encode information about the form and meaning of this word. Because the same set of features encodes patterns associated with all the words, the pattern of activation that is generated by the network as partial input is processed is a blend of the patterns associated with each lexical candidate that is consistent with that input. Given larger numbers of candidates consistent with a partial input, the distributed pattern generated by the network will become more and more distant from the pattern associated with the target word. This, in effect, is a form of lexical competition, best described in terms of interference between fully distributed patterns of lexical representation.

Marslen-Wilson and Warren (1994) presented evidence that, they argued, was inconsistent with competition operating via lateral inhibition. They created cross-spliced word sequences whose initial portion had been excised from another token of the same word (e.g., *jo(b) + (jo)b*, W1W1 sequence), from another existing word (e.g., *jo(g) + (jo)b*, W2W1 sequence), or from a nonword (e.g., *jo(d) + (jo)b*, N3W1 sequence). For the W2W1 (*jo(g)b*) and N3W1 (*jo(d)b*) sequences, formant transitions in the vowel provided misleading information about the place of articulation of the following consonant. Thus, these stimuli contained *subcategorical* phonetic mismatches (Streeter & Nigro, 1979; Whalen, 1984, 1991). They reasoned that if lexical candidates inhibit one another as predicted by TRACE, lexical decisions to words with subcategorical mismatches cross-spliced from words should be slower than lexical decisions to the words cross-spliced from nonwords. In TRACE, for W2W1 sequences, the initially activated competitor W2 (e.g., *jog*) inhibits the target W1 (e.g., *job*); in N3W1 sequences, this inhibition is substantially weaker because the nonword N3 (e.g., *jod*) only weakly supports both W2 and W1. Inhibition modifies the activation of words throughout processing. Thus, the degree to which the competitor W2 is activated affects the activation of the target W1 throughout the recognition process. Simulations with TRACE confirmed these predictions. The response probability for the target W1, calculated with a form of Luce's (1959) choice rule (roughly, a transformation of the activation of the target divided by the sum of all the other words' transformed activation), was substantially lower in W2W1 than in N3W1.

However, Marslen-Wilson and Warren (1994) found that the mean lexical-decision latencies to the W2W1 and N3W1 sequences did not differ from one another, whereas both were significantly longer than responses to the W1W1 sequences. This result was subsequently replicated by McQueen, Norris, and Cutler (1999). Marslen-Wilson and Warren argued that the absence of a difference between the W2W1 and N3W1 conditions provided strong evidence against models that incorporate lexical competition via lateral inhibition throughout the activation process.

More recently, Norris, McQueen, and Cutler (2000) have simulated the pattern of lexical decisions for stimuli with subcategorical mismatches using a small competition-activation model with lateral inhibition between active lexical candidates (the Merge model). They conducted simulations in which the word level was allowed to cycle through 15 iterations on each time slice, followed by a reset of lexical-activation levels, before the next slice was processed. Despite early and substantial differences in the activation of W1 given W2W1 and N3W1 (with W1's activation initially significantly depressed in the former case), the competition between W2 and W1 in W2W1 was resolved very quickly (in terms of input slices) in this network. Norris et al. showed that, with an appropriate decision threshold, lexical competition would be resolved before lexical decisions are initiated, yielding the observed latency pattern ($W1W1 < N3W1 \approx W2W1$). They argued that TRACE's failure in the Marslen-Wilson and Warren simulations was not due to lateral inhibition, but rather to lexical competition being resolved too slowly in TRACE. These conflicting simulations illustrate how difficult it is to distinguish among competing models without detailed information about the time course of activation of lexical competitors. The different patterns of activation predicted by models with and without lateral inhibition might occur too early in processing to be detected using lexical decisions.

Recently, a growing number of researchers, building upon work by Cooper (1974) and Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), have begun to use eye movements to explore questions about the time course of spoken-language comprehension, including the time course of spoken-word recognition in continuous speech. For example, Allopenna, Magnuson, and Tanenhaus (1998) had participants follow spoken instructions to pick up and move pictures using a computer mouse. On critical trials, participants saw displays containing items with similar names (e.g., *beaker*, *beetle*, and *speaker*, as well as an unrelated item). The probability of fixating each object as the target word unfolded over time was hypothesised to be closely linked to the activation of the lexical representation of this object (i.e., its name), under the assumption that the activation of the name of a picture influences the probability that a participant shifts attention to that picture and makes a saccadic eye

movement to fixate it. The minimum latency to plan and launch a saccade is estimated to be between 150 and 180 ms in simple tasks (e.g., Fischer, 1992; Saslow, 1967), whereas intersaccadic intervals in tasks such as visual search fall in the range of 200 to 300 ms (e.g., Viviani, 1990). Allopenna et al. found that the proportion of fixations to referents and competitors began to increase 200 to 300 ms after word onset, demonstrating that eye movements were sensitive to changes in lexical activation within the first 100 ms of the spoken word. Moreover, the probability of fixating each item over time as the target word was heard mapped closely onto predicted response probabilities from the TRACE model. Dahan, Magnuson, and Tanenhaus (2001) provided further support for the linking hypothesis between lexical activation and fixations in a set of experiments examining the time course of frequency effects on lexical activation. In conjunction with simulations using TRACE, these data provided strong support for models in which frequency has continuous, immediate effects on activation.

The Allopenna et al. and Dahan et al. studies demonstrate that the eye-tracking paradigm can provide detailed time-course information about lexical activation in continuous speech. It can provide a measure of competitors' activation over time when these competitors, along with the target, are visually displayed. Moreover, it can capture subtle time-course effects such as effects of frequency. Furthermore, the task (i.e., identifying the referent picture) requires participants to map the target word's auditory form onto semantic and visual information: This ensures that the task is specifically tapping into the process of word recognition. Thus, the paradigm appears well suited to measuring the target's activation over time in the three cross-splicing conditions studied by Marslen-Wilson and Warren (1994).

The present study had two goals. First, we used the eye-tracking paradigm to track the time course of lexical activation and hence to examine the time course of lexical competition. Eye movements to the picture associated with the target W1 were monitored in all three cross-splicing conditions, with the picture associated with the competitor W2 absent from the display (Experiment 1) or present (Experiment 2). Results from both experiments provided clear evidence for lexical competition. Second, we evaluated whether fixation patterns mapped onto predictions generated by TRACE.

EXPERIMENT 1

The goal of this experiment was to track the time course of lexical activation for cross-spliced target words. We monitored participants' eye movements to pictured objects as they heard a referent's name in each of

three splicing conditions: W1W1, W2W1, and N3W1. We hypothesised that the latency with which participants would make an eye movement to fixate the target picture (associated with W1) would reflect the target W1's lexical activation. In order to minimise the proportion of trials where participants were already fixating the target picture at the onset of the target word, participants were first instructed to point with the mouse cursor to one of the displayed distractor pictures (e.g., "Point to the bass"). As soon as the mouse cursor reached the picture, the critical instruction containing the target word was played (e.g., "now the net"). The purpose of this procedure was to draw participants' attention toward the distractor picture, thus away from the target picture, just before the target word was heard.

Method

Participants. Thirty students at the University of Rochester participated in this experiment and were paid \$7.50. All were native speakers of English.

Materials. Fifteen triplets composed of two real words and a nonword were selected (e.g., *net*, *neck*, **nep*). These items are listed in Appendix A. All the items were monosyllabic and ended with a stop consonant (labial /b/ or /p/, coronal /d/ or /t/, or velar /g/ or /k/). Within the triplet, one word was assigned the role of target (W1), and the other, the role of competitor (W2). The nonwords (N3) were constructed by changing the place of articulation of W2's final consonant. A number of constraints applied in the selection of these triplets. The primary constraint was that, by changing the place of articulation of the final stop consonant of a word, both a word and a nonword were generated. For the purposes of Experiment 1, the targets, and, for Experiment 2, both the targets and competitors, had to be picturable nouns (e.g., *net*, *neck*). The final constraint imposed on the nonwords was that they not correspond to the initial sequence of many real words, to minimise lexical activation that the nonword sequence could generate (e.g., the triplet *bat/back/*bap* was excluded because **bap* corresponds to the initial sequence of many words, such as *baptism* and *baptise*). To reach a reasonable number of experimental triplets, we had to relax these criteria for some items. In particular, we used the sequences *tat* and *hark* as nonwords, even though they are words but of extremely rare use. If participants were to treat them as words, this would cause the W2W1 and N3W1 conditions to become more similar, and hence go against our hypothesis that performance in these conditions should differ. Moreover, some nonword sequences matched the beginning of some other words, but only rare words (e.g., **nep* matches *Neptune* and *nepotism*).

Finally, the voicing feature of the stop consonant was kept constant for all three items for seven triplets (e.g., *net*, *neck*, **nep*), while the voicing feature of the target differed from that of the competitor and the nonword for the other eight triplets (e.g., *pit*, *pig*, **pib*). Because the crucial comparison was between the W2W1 and N3W1 conditions, both W2 and N3 shared the same voicing. As the results will show, these two sets of triplets yielded similar patterns of results.

In order to generate cross-spliced stimuli, each item of the 15 triplets was recorded by a male native speaker of American English in a sound-proof room, sampling at 22050 Hz with 16-bit resolution. Each item was embedded in the sentence used as the critical instruction in the experiment (e.g., “now the *net/neck/*nep*”). Each sentence was then edited. The final stop consonants of the W1 items were spliced onto the initial portion (up to the end of the vowel) of another token of W1 (e.g., *now the ne(t) + t*), or of a token of W2 (e.g., *now the ne(ck) + t*), or of a token of N3 (e.g., *now the ne(p) + t*). This procedure generated three versions for each of the 15 experimental target words. On average, the duration of the *now the* part of the instruction was 461 ms for the W1W1 stimuli, 468 ms for the W2W1 stimuli, and 452 ms for the N3W1 stimuli; the duration of the target word up to the end of the vowel (i.e., before the splicing point) was 376 ms, 378 ms, and 383 ms for the W1W1, W2W1, and N3W1 stimuli, respectively. The average duration of the last consonant was 206 ms.¹

¹ To test that coarticulatory cues in W2W1 sequences were not stronger than those in N3W1 sequences, we conducted a forced-choice phonetic decision experiment (see McQueen et al., 1999, Experiment 6). Listeners heard only the vowels excised from the 45 experimental items (e.g., [ɛ] from *net*, *neck*, and **nep*), and had to indicate, for each vowel, what the following segment had been by choosing among two response alternatives. The vowel plus the correct following consonant formed an existing word in 11 items (5 for W1, 2 for W2, and 4 for N3). McQueen et al. controlled for possible lexical biases in the consonant choice by having both consonant alternatives (the target and the distractor) forming either a word or a nonword. However, this procedure had the disadvantage of including consonant distractors that were never targets on other trials. We neutralised lexical bias by equating the number of (correct and incorrect) lexical alternatives between the W2 and N3 items. Only stop consonants were used as response alternatives, and six different pairings were used, with the following frequency: K-P (14), K-T (7), T-P (7), B-G (10), B-D (5), D-T (2). Each of the 45 vowels was presented three times; the 135 trials were presented in random order. Twelve participants were tested. On average, the percentage of correct responses was 78%, 69%, and 71%, for W1, W2, and N3 items, respectively, and did not vary significantly ($F_1(2,22) = 3.09$, $F_2(2,28) = 1.08$). Crucially, W2 items did not yield more correct responses than N3 items ($F_1(1,11) = 0.43$; $F_2(1,14) = 0.09$), indicating that coarticulatory cues to the following consonant in the vowel were equally strong in both sets of items. Any difference in performance when the items are fully presented cannot be attributed to differences in amount of coarticulatory information in the W2W1 and N3W1 sequences.

For each of the 15 experimental target words, three distractor words were selected. They were all picturable monosyllabic words. One of them began with the same consonant as the target word (e.g., for the target word *net*, the distractor was *nurse*). This was done to prevent participants from identifying the target picture on the basis of the initial consonant alone, before hearing the crucial subcategorical mismatch carried by the vowel. The other two distractor words were selected so that their names were phonologically highly dissimilar from the target. Items used in the experimental trials are listed in Appendix B.

In addition to the 15 experimental trials, 15 filler trials were constructed. For each trial, four picturable monosyllabic words were selected. For ten fillers, two picture names overlapped at onset (e.g., *bed* and *bell*) and neither of them was the target. This aspect of the fillers was especially designed for Experiment 2, where both W1 and W2 were visually present in the display, to prevent participants from developing expectations that pictures with phonologically similar names were likely to be targets. For the five other fillers, the four pictures' names were phonologically dissimilar. For eight of the 15 fillers, the trial structure was identical to that of the experimental trials: Participants were first instructed to point to a picture, then to another picture (e.g., "Point to the star. Now the goat. Click on it and put it below the diamond."). For seven of the fillers, participants were instructed to click on the picture they initially pointed to (e.g., "Point to the frog. Click on it and put it above the triangle.").

The spoken instructions for the filler trials were recorded by the same speaker as those for the experimental trials, during the same session. Subcategorical mismatches on ten instructions of the filler trials were created to prevent critical trials from being identified as those with cross-spliced stimuli (e.g., the onset and nucleus of the target word *cup* was cross-spliced with the final consonant of its counterpart *cut*).

The 120 pictures ([15 experimental + 15 filler] trials \times 4 pictures) were all black and white line-drawings. They were selected from the Snodgrass and Vanderwart (1980) and the Cycowicz, Friedman, Rothstein, and Snodgrass (1997) picture sets, as well as from children's picture dictionaries and a commercially available clip-art database.

Procedure. Prior to the eye-tracking experiment itself, participants were first exposed to each of the 120 pictures and familiarised with each intended name. This pre-exposure ensured that each picture was clearly identified and labelled as intended. Each picture was presented on a computer screen along with its printed name. Participants were free to inspect the picture as long as necessary, and moved to the next picture by pressing the keyboard's space bar.

The eye-tracking part of the study immediately followed. Participants were seated at a comfortable distance from the computer screen. Eye gaze was monitored using an Applied Scientific Laboratories head-mounted eye-tracker (model E 5000). A scene camera was aligned with the participant's line of sight. A calibration procedure allowed software to superimpose crosshairs showing the point of gaze on a HI-8 video tape record of the scene camera. The scene camera sampled at a rate of 30 frames per second, and each frame was stamped with a time code. Auditory stimuli were played to the participant through headphones and simultaneously to the HI-8 VCR, providing an audio record of each trial.

The structure of each trial was as follows: First, a 5×5 grid with a centred cross appeared on the screen, and participants were told to click on the cross. This allowed the experimenter to check that calibration accuracy was acceptable, as participants briefly fixated the cross before clicking on it. Then four line-drawings and four coloured geometric shapes appeared on specific cells of the grid. As the pictures appeared on the screen, the first spoken instruction started, asking participants to point to one of the distractor pictures using the computer-mouse cursor (e.g., "Point to the bass"). As soon as the cursor reached the distractor picture, the second and critical instruction was played, instructing participants to point to the target picture (e.g., "now the net"). Participants were then told to move the target picture above or below one of the geometric shapes (e.g., "Click on it and put it above the circle"). Once this was accomplished, the next trial began. On some filler trials, participants were instructed to click on and move the first picture they pointed to (e.g., "Point to the key. Click on it, and put it below the square"). This was intended to ensure that people directed their attention to the first picture. The positions of the geometric shapes were fixed from one trial to the other. The position of each picture was randomised for each participant and each trial. Five fillers were presented at the beginning of the session, to familiarise participants with the task.

Three lists were constructed by varying which of the three versions of each target word was presented (W1W1, W2W1, or N3W1). Within each list, five critical items were assigned to each condition. Ten participants were randomly assigned to each list. For each list, three random orders were created; approximately the same number of participants were assigned to each random order.

The data were collected from the videotape records using an editing VCR with frame-by-frame controls and synchronised video and audio channels. Coders used the crosshairs generated by the eye tracker to establish, for each experimental trial, which of the four pictures or the cross was fixated at each time frame (see Dahan, Swingley, Tanenhaus, & Magnuson [2000], for full details on the coding procedure).

Results

Analysis of latency. We measured, for each participant and each trial, the latency (from target onset) with which the participants fixated the target picture immediately before clicking on it with the mouse. This mouse-cursor movement was taken as an indication that participants had recognised the target word. Despite our efforts to draw participants' attention away from the target picture, participants fixated this picture at target onset on 18.1% of the trials (the non-mentioned distractor pictures were fixated on 17% of the trials). However, on most such trials, participants fixated another picture before returning to the target; on only a few trials did they keep fixating the target (21 out of 450, 4.7%). These trials were excluded from subsequent analyses. For seven participants, one additional trial was missing because of technical failures.

Table 1 presents the mean latency for each splicing condition, as a function of the voicing status within the triplet (same voicing, as in *net/neck/*nep*, or different voicing, as in *pit/pig/*pib*). The mean latency to fixate the target picture was 638 ms in W1W1, 851 ms in W2W1, and 673 ms in N3W1. A two-way ANOVA (splicing condition \times voicing status) revealed a significant effect of splicing condition ($F_1(2, 58) = 13.7, p < .001, MSE = 57092.5; F_2(2, 26) = 12.5, p < .001, MSE = 14154.3$), no effect of voicing, and no interaction with splicing condition. Newman-Keuls tests indicated that the latency was significantly slower in W2W1 than in the W1W1 and N3W1 conditions, with no significant difference between W1W1 and N3W1 (with $\alpha = .05$). The mean latency in W1W1 is similar to that found for non-cross-spliced targets of similar lexical frequency (Dahan et al., 2001).²

The latency analysis revealed a significant difference between W2W1 and N3W1, as well as between W1W1 and W2W1. By contrast, N3W1 and W1W1 did not differ significantly, although latencies were numerically slower in N3W1. Participants were slower at recognising the target word and identifying the referent picture when mismatching coarticulatory information in the target word's vowel matched a real word (as in the W2W1 sequences) than when this coarticulatory information did not match an existing word (as in the N3W1 sequences). Very little effect of

² Additional analyses were conducted after excluding trials for which the latency was less than 200 ms or more than 1500 ms. Fixations occurring before 200 ms were likely to have been programmed before the onset of the target word, and not resulting from processing the target word. Trials where participants took more than 1500 ms to fixate the target were treated as outliers. In total, 38 out of the 422 remaining trials were excluded (9.0%, 11, 11, and 16 in W1W1, W2W1, and N3W1, respectively). Analyses yielded the same pattern of results as found in the full data set.

TABLE 1

Experiment 1. Mean latency (in ms) and standard errors (between parentheses) for each splicing condition as a function of the voicing status within the triplet (same-voice vs. different-voice)

	<i>Same-voice</i>		<i>Different-voice</i>	
	<i>Example</i>	<i>Latency</i>	<i>Example</i>	<i>Latency</i>
W1W1	ne(t)t	639 (44)	pi(t)t	636 (40)
W2W1	ne(ck)t	834 (39)	pi(g)t	867 (59)
N3W1	ne(p)t	712 (51)	pi(b)t	634 (57)

mismatching information was observed when the coarticulatory information did not match an existing word.

Analysis of fixations over time. We computed the proportions (across participants) of fixations to the target picture over time, for each of the three splicing conditions. For each participant and each trial, we established which of the four pictures or the cross was fixated at each time frame, beginning at the onset of the target word. The proportion of fixations to each picture at each time frame was then computed for each participant, and these proportions were averaged across participants. Figure 1 presents the proportions of fixations to the target picture over time for each splicing condition. Fixations between conditions were comparable until about 600 ms after target onset, where the fixations in W2W1 started diverging from those in W1W1 and N3W1. Recall that the duration of the pre-splice fragment was about 400 ms, with coarticulatory cues being presumably strongest in the late portion of the vowel. Given a 200-ms delay to program and launch an eye movement, fixations occurring around 600 ms are likely to result from the processing of the coarticulatory information. When this information matched an existing word, as in W2W1, fixations to the target were considerably delayed; when this information did not match a word, as in N3W1, no such delay was observed. Difference in fixations between W2W1 and N3W1 extended until about 1200 ms after target onset. A two-way ANOVA (splicing condition \times voicing status) on mean fixation proportions over a time window extending from 600 to 1200 ms after target onset revealed a significant effect of splicing condition ($F_1(2, 58) = 12.55, p < .0001, MSE = 0.0376; F_2(2, 26) = 15.06, p < .0001, MSE = 0.008$), no main effect of voicing, and no interaction with splicing condition. Planned comparisons revealed a significant difference in target fixations over the 600- to 1200-ms window between the N3W1 and W2W1 conditions ($t_1(29) = 3.63, p < .005, t_2(14) = 3.33, p < .01$) and between the W1W1 and W2W1 conditions

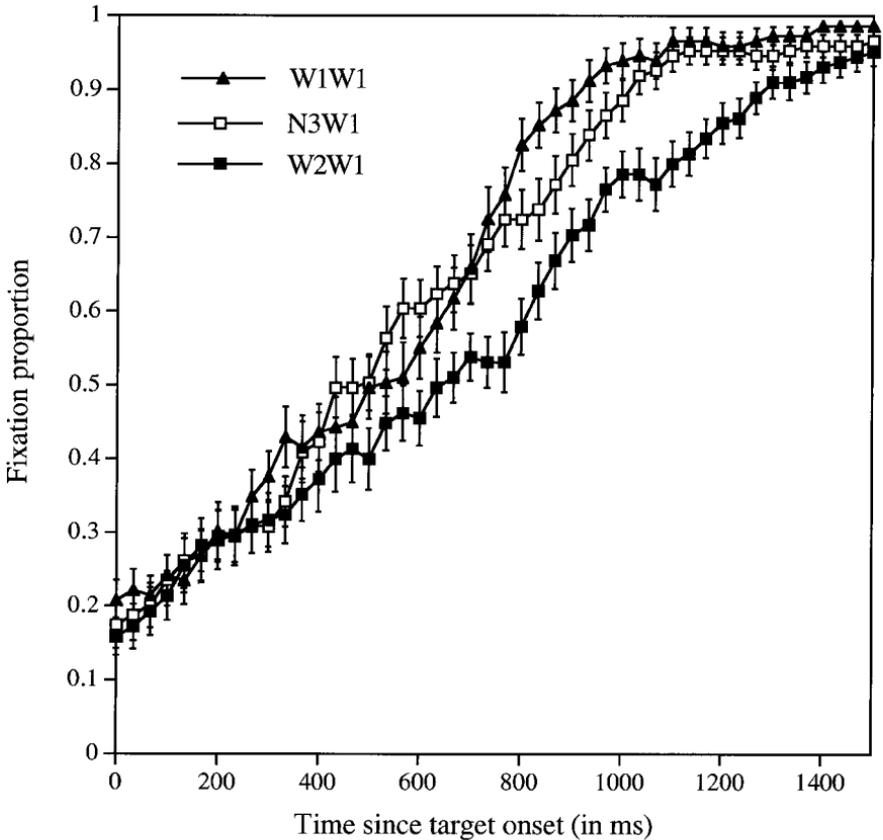


Figure 1. Experiment 1: Proportion of fixations to the target picture (W1) over time for each splicing condition (W1W1 [filled triangles], W2W1 [filled squares], N3W1 [empty squares]). Bars indicate standard errors.

($t_1(29) = 5.56, p < .0001, t_2(14) = 5.42, p < .0001$), but no significant difference between the W1W1 and N3W1 conditions.

The fixation analysis revealed a competition effect extending roughly from 600 to 1200 ms after target onset. Given that the mean duration of the target word was 585 ms and even after adding a 200 ms delay for programming an eye movement, this suggests that the competition between W2 and W1 may persist for several hundred ms after the disambiguating information is encountered. However, it is possible that the eye-movement data exaggerate how long it takes for lexical competition to resolve because on some trials, participants made multiple eye movements before fixating the target, and each new fixation involved some motor delay. In order to eliminate effects due to multiple fixations, we conducted a subanalysis that included only those trials on which participants made a

single fixation from the onset of the target word, and in which that fixation was to the target. (That fixation always began from any picture except the target picture.) This analysis included 23.5% of the trials. Figure 2 presents fixation proportions to the target picture over time in each splicing condition. As is apparent from the figure, evidence of lexical competition (i.e., difference between the W2W1 and N3W1 conditions) extended until about 1000 ms after target-word onset. Recall that, on the main analysis, competition effects were observed as early as 200 ms after the offset of the pre-splice portion (at 600 ms, with a mean pre-splice portion of 380 ms), confirming that 200 ms is a reasonable estimate of the delay for observing effects of processing and integrating the spoken input in the eye-tracking task. Moreover, the post-splice consonant was 206 ms long on average. If lexical competition were to resolve as soon as disambiguating information

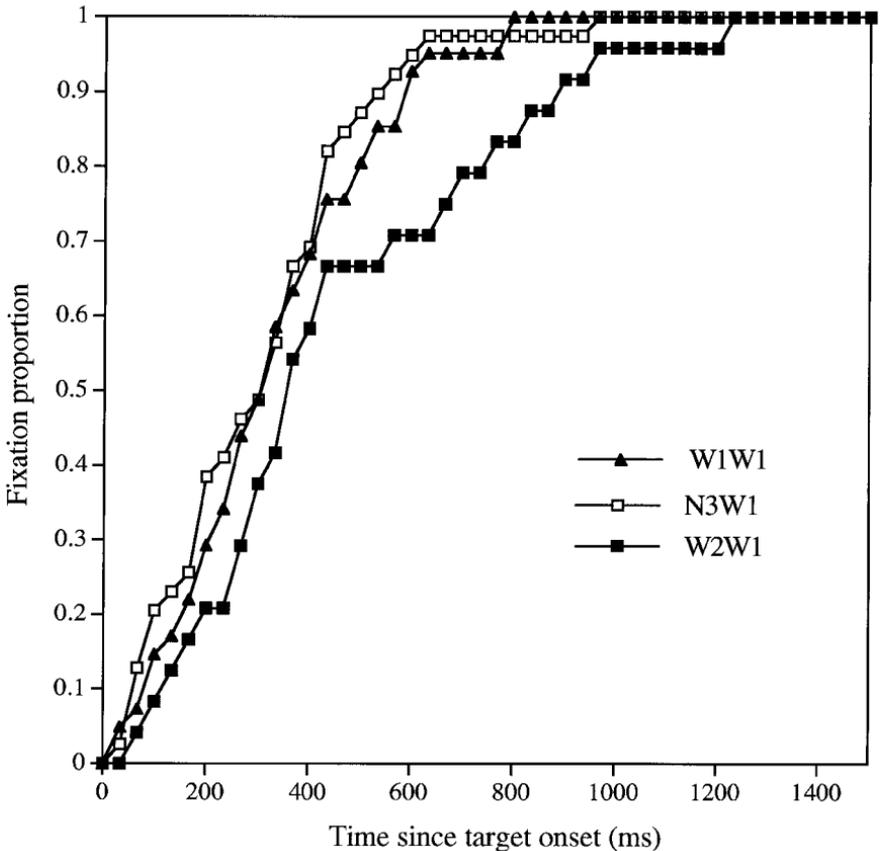


Figure 2. Experiment 1: Proportion of fixations to the target picture (W1) over time for each splicing condition (W1W1 [filled triangles], W2W1 [filled squares], N3W1 [empty squares]) on single-fixation trials (see text).

has been heard and processed, evidence for such resolution would be expected around 800 ms after target onset. Although this analysis should be interpreted with caution (no statistical analyses were conducted, due to the small number of data points per subject), it suggests that lexical competition continues for several hundred ms beyond the point where the disambiguating information has been processed.

Discussion

The latency with which participants fixated the target picture was slower when mismatching information in the target word matched a competitor word than when it did not. The latency analysis showed a small (nonsignificant) difference between the W1W1 and N3W1 conditions, whereas the difference between the W1W1 and W2W1 conditions was substantial. This suggests that lexical access is much more disrupted when mismatching coarticulatory information matches a word than when it does not. Eye movements to the target picture thus confirm the general prediction made by models that assume lexical competition throughout the recognition process. Early in the W2W1 sequence, the competitor W2 becomes highly active and competes with the target W1. The recognition of the target word is thus delayed. In contrast, early in the N3W1 sequence, W2 is only weakly active, so its activation has a much smaller effect on the recognition of W1. The eye-movement analyses capture the time-course aspect of the competition effect: Fixations to the target picture in the W2W1 and N3W1 conditions remain fairly similar up to 600 ms after target onset; at this point, the proportion of target fixations in W2W1 is delayed compared to the other conditions, as a result of the strong activation of the competitor (W2).

EXPERIMENT 2

Experiment 1 demonstrated that the recognition of the target word is more delayed when it contains mismatching coarticulatory information that matches an existing competitor word than when the mismatching information does not favour a potential lexical competitor. Experiment 2 was designed to assess more directly the competitor's activation by measuring fixations to this competitor over time in each of the splicing conditions. Furthermore, fixations to the target and its competitor over time provide the basis for a direct comparison between these fixation patterns and the activation patterns generated by a model that incorporates lexical competition. In order to obtain time-course data on the activation of both the target and the competitor in each of the splicing conditions, we presented the competitor picture along with the target picture and two distractors.

Method

Participants. Thirty students at the University of Rochester participated in the experiment and were paid \$7.50. All were native speakers of English. None of them had participated in Experiment 1.

Materials. The materials were identical to those used in Experiment 1, with the exception of the displayed pictures. Here, the competitor picture was displayed along with the target picture and two unrelated distractors (e.g., the picture of a net, a neck, a bass, and a deer). The list of the pictures is presented in Appendix B.

Procedure. The procedure was identical to that used in Experiment 1. Three lists were constructed, by varying which of the three versions of each experimental target word was presented (W1W1, W2W1, or N3W1). Within each list, five critical items were assigned to each condition. Ten participants were randomly assigned to each list. For each list, three random orders were created; approximately the same number of participants were assigned to each random order.

Results and discussion

Because of technical failures, five trials were missing; in addition, for nine trials (2%), participants erroneously clicked on the competitor picture. These trials were excluded from the analyses. Figure 3 presents the fixation proportions to the target W1 (upper panel) and its competitor W2 (lower panel), for each splicing condition. Fixations to the target over time indicated a fast rise in W1W1, separating from the other conditions shortly after 600 ms; the target fixations rose more slowly in N3W1, and slowest in W2W1. Fixations to the competitor W2 revealed a complementary picture. The competitor picture was fixated most in W2W1, where coarticulatory information in the vowel matches the competitor's name, intermediate in N3W1, where coarticulatory information weakly matches both W1 and W2, and least in W1W1, where coarticulatory information favours W1. In the latter condition, fixations to target and competitor increased in parallel until shortly after 600 ms, where competitor fixations began to decrease, and target fixations, to increase. Fixations at this point thus reflect the use of coarticulatory information in the vowel supporting the target over the competitor.

We computed the mean proportion of fixations to the target and competitor for each splicing condition over the 600- to 1200-ms time window. On average, the target was fixated most given W1W1, substantially less given N3W1, and least given W2W1 (70%, 49%, and 42%, respectively); a complementary pattern was observed for the

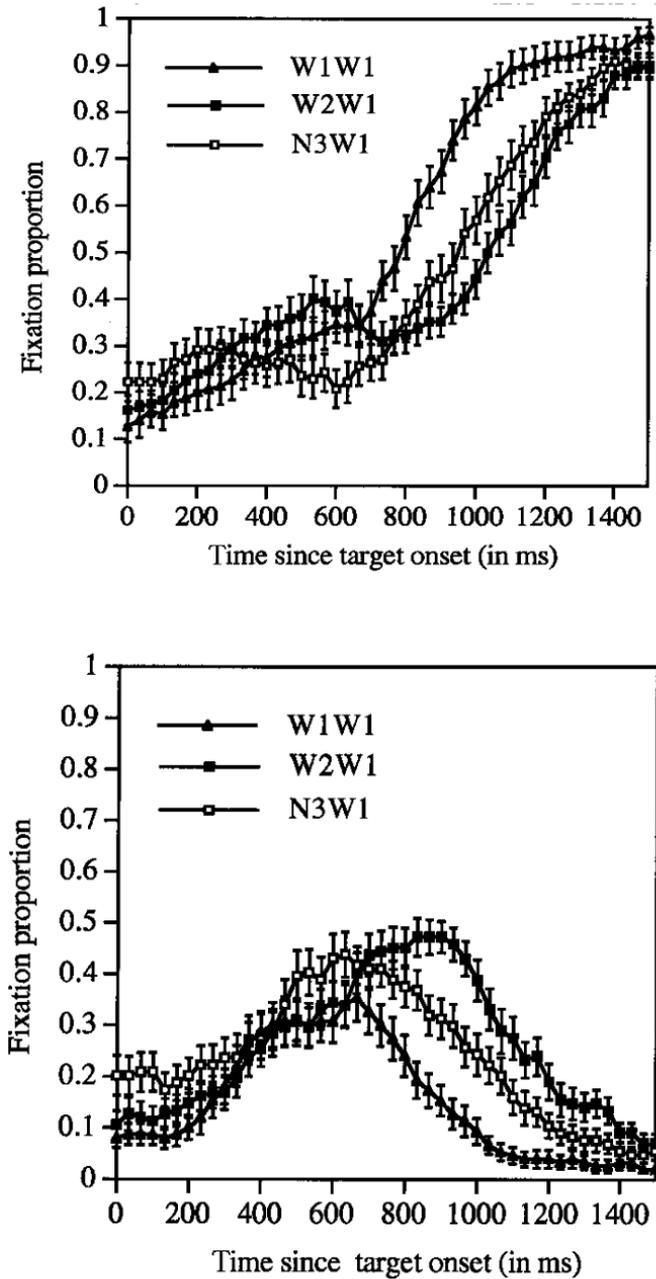


Figure 3. Experiment 2: Proportion of fixations to the target picture (W1) (upper panel) and the competitor picture (W2) (lower panel) over time, for each splicing condition (W1W1 [filled triangles], W2W1 [filled squares], N3W1 [empty squares]). Bars indicate standard errors.

competitor (15%, 29%, and 39%, respectively). A three-way ANOVA (picture [target or competitor] \times splicing condition \times voicing status) revealed a main effect of picture ($F_1(1, 28) = 56.6, p < .0001, MSE = .0612; F_2(1, 13) = 16.93, p < .001, MSE = .0607$) and a significant interaction between splicing condition and picture ($F_1(2, 56) = 26.65, p < .0001, MSE = .0603; F_2(2, 26) = 23.26, p < .0001, MSE = .0159$), reflecting the complementary pattern found for the target and competitor fixations across the splicing conditions. (The data from one subject were excluded from the subject analyses because of missing data causing empty cells.) No other main effect or interaction reached significance. A similar pattern of results (although marginally significant by items) was found when the analysis was restricted to the two crucial conditions (W2W1 and N3W1) and averaged across voicing status. There was a main effect of picture ($F_1(1, 29) = 28.86, p < .0001, MSE = .0320; F_2(1, 14) = 3.3, p = .09, MSE = .0467$) and an interaction between picture and splicing condition ($F_1(1, 29) = 5.32, p < .05, MSE = .0225; F_2(1, 14) = 2.8, p = .12, MSE = .0275$). Pairwise comparisons revealed a significant difference between W2W1 and N3W1 only on competitor fixations ($t_1(29) = 2.52, p < .05, t_2(14) = 1.99, p = .06$).³

Fixations to both the target and the competitor pictures over time revealed differences between the splicing conditions. Given W1W1, the target (W1) and its competitor (W2) were both equally activated until the sensory input, presumably the coarticulatory information in the vowel, favoured W1 over W2. While the input was consistent with both W1 and W2, these items were equally likely to be fixated. This is consistent with “cohort” effects previously demonstrated with the eye-tracking paradigm (Allopenna et al., 1998; Dahan et al., 2000; Tanenhaus et al., 1995). When coarticulatory information was encountered in the vowel given W1W1 or W2W1, fixations to the mismatching item decreased. Participants’ sensitivity to coarticulatory information was revealed by the contrastive patterns for W1 and W2 early on given W1W1 and W2W1. Eye movements thus capture sensitivity to fine-grained information in the sensory input. Furthermore, the lesser decrease in fixations to W1 given

³ We also conducted an analysis on the latency in fixating the target picture. On average, the latency was 819 ms in W1W1 (standard error = 33 ms), 1110 ms in W2W1 (standard error = 42 ms), and 1052 ms in N3W1 (standard error = 43 ms). A two-way (splicing condition \times voicing status) ANOVA showed a main effect of condition ($F_1(2,56) = 14.9, p < .0001, MSE = 91887; F_2(2,26) = 17.27, p < .0001, MSE = 17379$), with no main effect of voicing and no interaction. Pairwise comparisons revealed significant differences between W1W1 and both W2W1 and N3W1, but not between W2W1 and N3W1. Note that both mean latencies and variability were noticeably greater than in Experiment 1, certainly due to the eye movements made to the competitor picture before fixating the target picture. Nevertheless, there was a trend toward slower latencies in fixating the target picture in W2W1 than in N3W1, confirmed in the fixation-proportion analysis.

N3W1 compared to W2W1 revealed lexical competition. The effect of mismatch was substantially less when the mismatching information did not favour any lexical alternative.

SIMULATIONS OF THE EYE-MOVEMENT DATA WITH THE TRACE MODEL

In order to test whether the eye-movement data are consistent with predictions from a model that incorporates lexical competition via lateral inhibition, we simulated the fixations to the target and competitor by transforming activation generated by TRACE into fixation probabilities. We used the publicly available TRACE implementation (<ftp://ftp.crl.ucsd.edu/pub/neuralnets>) with the standard parameter set reported in McClelland and Elman (1986).⁴ The lexicon was augmented to 257 words to include the closest possible transcriptions of our stimuli given TRACE's limited phoneme set. The transcriptions of our stimuli are presented in Appendix A.

TRACE provides a coarse approximation of coarticulation by spreading features from a segment forward and backward six slices, such that each segment spans 11 cycles. Segments overlap because adjacent phoneme centres are six cycles apart. Features spread with a triangular function, such that they peak at each phoneme centre, and decrease gradually forward and backward from the peak. In our simulations, words were preceded by 6 cycles of silence; thus, the onset of each word was at cycle 7, with phoneme centres at cycle 12, 18, and 24. In order to generate W2W1 and N3W1 cross-spliced inputs, the cross-splicing point was chosen to follow the time slice immediately preceding the centre of the last consonant. Input stimuli were presented to TRACE, one at a time. All lexical items were allowed to compete.

Figure 4 shows the raw TRACE activation over time for W1 and W2 nodes as W1W1, W2W1, and N3W1 inputs were presented, averaged over the 15 items (upper panel). These activation patterns were converted into predicted fixation probabilities over time using a variant of the Luce choice rule applied to the four visually present alternatives, i.e., W1, W2, and two phonologically unrelated distractors (for a more detailed explanation of the issues involved in mapping activation onto fixation proportions, see Allopenna et al., 1998; Dahan et al., 2001; Magnuson, Tanenhaus, Aslin, & Dahan, submitted). Fixation probabilities are shown in the lower panel of Figure 4. The fixation probability for W1 rises fastest given W1W1,

⁴ The default parameter set in the distribution version differs from that reported by McClelland and Elman (1986). In the original paper, all features were set to spread forward and backward six slices, whereas in the distribution code, different features spread different numbers of slices.

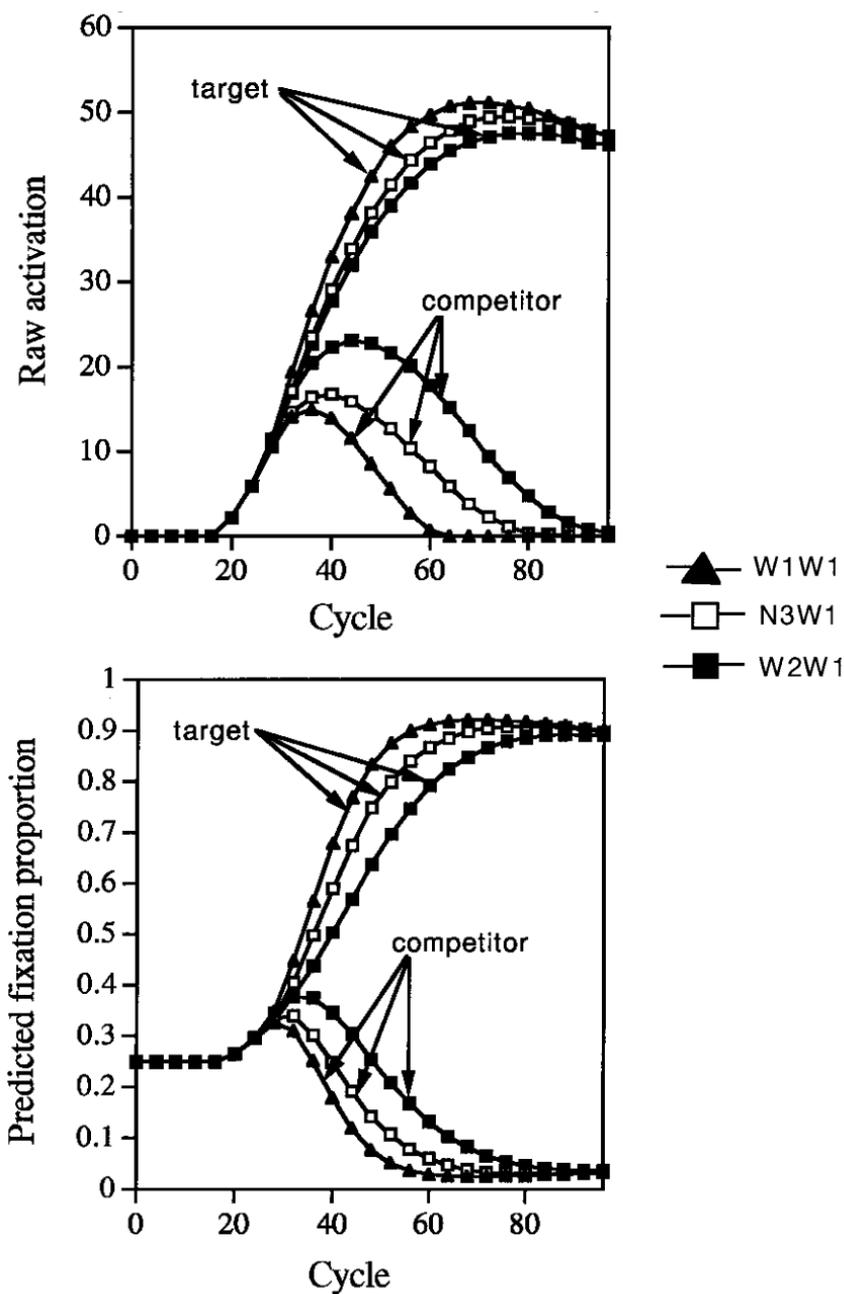


Figure 4. Simulations: Raw TRACE activation (upper panel) and predicted fixations (lower panel) for the target (W1) and the competitor (W2), over cycles, for each splicing condition (W1W1 [filled triangles], W2W1 [filled squares], N3W1 [empty squares]).

somewhat slower given N3W1, and slowest given W2W1. Note that W1 fixations reach a high probability in all conditions, even when the input in the vowel favours its competitor W2, as in W2W1 sequences. Activation levels are thus modulated by information occurring later in the input, counteracting earlier information. The model predicts more fixations to W2 given W2W1, less given N3W1, and the least given W1W1. The predicted fixations closely mirror the human data shown in Figure 3.

These simulations diverge from the TRACE simulations reported by Marslen-Wilson and Warren (1994) in important ways. Their simulations did predict a lower probability of recognising the target (W1) given W2W1 than given N3W1 or W1W1, in agreement with our human data and simulations. However, the probability of recognising W1 given W2W1 was only temporarily delayed in our simulations, while it remained very low even after 80 cycles in Marslen-Wilson and Warren's. Furthermore, they simulated lexical activation to cross-spliced nonword sequences (W2N1, e.g., *smo(g)b*), and found that the probability of recognising W2 given W2N1 reached roughly the same level as the probability of recognising W1 given W1W1 (see their Figures 12 and 13). Translated into lexical-decision judgements, these probabilities predict high error rates in lexical decisions given W2N1, because the nonword sequences W2N1 receive as much "word" support as the word sequences (W1W1 or N3W1). However, lexical decisions to the nonword sequences W2N1 had a very low error rate, both in the Marslen-Wilson and Warren (1994) and the McQueen et al. (1999) studies. Marslen-Wilson and Warren attributed the discrepancies between the human data and their simulations to TRACE's architecture (in particular, the presence of lateral inhibition between word nodes and absence of bottom-up inhibition). Because our simulations of the same conditions did not show the patterns reported by Marslen-Wilson and Warren, this suggests that these patterns resulted more from parameter and/or stimulus choices than from the principles underlying TRACE's architecture.⁵

⁵ We have not been able to uncover the source of the discrepancies between our simulations and those reported by Marslen-Wilson and Warren (1994). However, it is clear that the pattern we report represents typical TRACE behaviour. It holds for splicing at multiple locations and three different parameter sets (those reported in McClelland & Elman [1986], the default parameter settings in the Unix distribution of the TRACE simulator, and the parameters described by Frauenfelder and Peeters [1998]). Similar results hold across different variants of the Luce choice rule with a range of values for the *k* parameter (see Magnuson, Dahan, and Tanenhaus, 2001, for details), different sets of stimuli (the CVC analogs to our stimuli, every possible stimulus set in the lexicon used by Marslen-Wilson and Warren, and the specific set of five triplets used by Marslen-Wilson and Warren [league/lead/*leab, pug/pub/*pud, slab/slag/*slad, stab/stag/*stad, and shrub/shrug/*shrud; Warren, personal communication]), and two different implementations of TRACE (the Unix

In summary, our TRACE simulations mirror the eye-movement data, with delayed fixation probabilities over time for the target (W1) in W2W1 compared to N3W1, caused by a high fixation probability for W2.

GENERAL DISCUSSION

The present study examined lexical competition when mismatching coarticulatory cues in a vowel (i.e., inconsistent with the actual identity of the following consonant) match another existing word and when these cues do not match an existing word. Earlier studies (Marslen-Wilson & Warren, 1994; McQueen et al., 1999) collected lexical-decision latencies to cross-spliced word sequences containing mismatching coarticulatory cues and found no effect of the lexical status of the conflicting cues. Marslen-Wilson and Warren (1994) interpreted this result as evidence against lateral inhibition between activated word units, as instantiated in the TRACE and Shortlist models. However, Norris et al. (2000) showed that the absence of a lexical effect between W2W1 and N3W1 is compatible with a model incorporating lateral inhibition if lexical competition is resolved before lexical-decision responses are generated. On their account, such lexical competition takes place but is resolved before activation reaches a sufficient threshold to generate a lexical decision.

The present eye-movement data showed that the recognition of W1 is delayed given W2W1, compared to N3W1, demonstrating clear lexical competition between the competitor (W2) and the target (W1), regardless of whether W2 was visually present or not. Using the linking hypothesis developed by Allopenna et al. (1998) and Dahan et al. (2001), we showed that lexical activation generated from TRACE mirror the eye-movement data quite closely. Predicted fixations to the competitor W2 reach a higher level and fixations to the target W1 are more delayed in W2W1 than in N3W1. Importantly, the subcategorical mismatch affects activation only *temporarily*. The target's activation reached the same maximum value in all splicing conditions, suggesting that the target will always be ultimately recognised. This contrasts with the simulations reported in Marslen-Wilson

implementation, and the Macintosh implementation, "MacTRACE", provided to us by Paul Warren). Although we obtained stronger competition effects given W2W1 for some items than for others (e.g., longer words, especially the CCVCs used by Marslen-Wilson and Warren), the patterns more closely resembled the results shown in Figure 4 than the pattern reported by Marslen-Wilson and Warren. In short, we were unable to replicate the pattern reported in Marslen-Wilson and Warren (1994) with any combination of these variables. Thus, it seems safe to conclude that whereas some combination of stimuli and parameters may result in simulations with the pattern presented in Marslen-Wilson and Warren (1994), the simulations that we report in this article are more representative of the results that arise from the TRACE architecture.

and Warren (1994), where the ultimate response probabilities favoured the pre-splice portion of the input (e.g., W2's activation given W2N1 was at least as high as W1's activation given W1W1), leading researchers to incriminate various aspects of TRACE's architecture, such as lateral inhibition and lack of bottom-up mismatch (Marslen-Wilson & Warren, 1994) or lack of "optimisation" at every input slice (Cutler, Norris, & McQueen, 2000; Norris et al., 2000). Although the source of the discrepancies between the present simulations and those reported by Marslen-Wilson and Warren remains unclear, the present results indicate that Marslen-Wilson and Warren's simulations cannot be considered decisive evidence against the architectural assumptions of TRACE (see discussion in footnote 5).

Both the eye-movement data and the TRACE simulations showed an effect of the lexical status of the pre-splice portion of the sequence. Target recognition, as assessed by fixations, was more delayed when the pre-splice portion of the sequence came from another word (as in W2W1) than when it did not (as in N3W1). However, lexical-decision studies (Marslen-Wilson & Warren, 1994; McQueen et al., 1999) showed no difference in response latencies between these conditions. Why do lexical-decision latencies fail to show a difference between the W2W1 and N3W1 conditions, while we find such a strong difference in the eye-movement data? McQueen et al. (1999) argued that the lexical-decision data show no difference between the W2W1 and N3W1 conditions because the lexical competition between W2 and W1 is quickly resolved and W1 dominates the activation pattern, "leaving no trace of that competition process in the [lexical-decision] responses made to the cross-spliced words" (page 1385). According to Norris and colleagues (Cutler et al., 2000; Norris et al., 2000), the dynamics of word-node activation in Merge (i.e., optimisation by letting the network cycle 15 times and resetting word nodes at each time slice) are required to account for the lexical-decision data. In their simulations, the competition between W2 and W1 is resolved by the end of the final consonant of the cross-spliced stimuli and W1's activation reaches a response threshold with the same delay in the W2W1 and N3W1 conditions. However, our eye-movement data, and in particular the subanalysis conducted on Experiment 1 (see Figure 2), suggest that effects of lexical competition in cross-spliced stimuli extend well beyond the target-word offset, that is, until after processing of the post-splice consonant is likely to have completed. Even when considering only single-fixation trials, the delay with which participants fixated the target picture was longer when the pre-splice portion of the target word matched a competing word than when it did not, and this effect extended about 200 ms after the post-splice consonant had been heard and processed. Thus, effects of competition can be observable for some time after sufficient input to resolve ambiguity has been heard

and processed. This apparently delayed resolution of lexical competition, with respect to the time course with which spoken input becomes available, seems more compatible with a model like TRACE, where a word's activation at any particular point in time depends on both its activation at the previous time step and the current input, resulting in gradual increases and decreases of word activation, than with a model like Merge (or Shortlist), where multiple activation cycles and a reset of word nodes result in an optimal interpretation of the input at each time slice. Further evidence suggesting that the impact of speech information on the state of the lexical system is not immediate is provided by Zwitserlood and Schriefers (1995).

If competition effects do extend several hundred ms after disambiguating information has been heard, why aren't these effects reflected in lexical-decision latencies? A possible explanation for the absence of competition effects in the lexical-decision responses, as suggested by Norris and colleagues, is that participants responded after lexical competition had resolved. Mean reaction times in the McQueen et al. study were about 470 ms after target offset in the W2W1 and N3W1 conditions. If processing and motor-response delays in lexical decisions can be as short as 200 ms, responses could have been generated after full competition resolution. Note that on this account, there is no need to assume optimisation in competition resolution. However, this apparent lack of sensitivity of the lexical-decision task to lexical competition is at odds with other evidence that lexical-decision responses to a target word are influenced by the activation of its competitors. For instance, Luce and Pisoni (1998) showed that lexical-decision latencies to words in high-density neighbourhoods (i.e., words with many phonologically similar competitors) are slower than latencies to words in low-density neighborhoods. Why would response latencies to cross-spliced words, for which the initial activation of the pre-splice competitors is certainly very high, fail to show a competition effect? If anything, one would expect to see an especially strong competition effect with such stimuli.

As an alternative account for the absence of lexical-competition effects in the lexical-decision data, we propose that lexical-competition effects are masked in the mean lexical-decision latencies because the activation of competitors as well as targets influences lexical-decision responses. The lexical-decision task does not require correct identification of the intended target. Participants may also respond 'yes' in response to the high activation of a competitor item. More specifically, some 'yes' responses to W2W1 sequences would be triggered by W2's early strong activation, while the impact of competitors' activation would be small on responses to N3W1 sequences. The mean lexical-decision latency in the W2W1 condition would thus include a range of latencies, from early responses

to the initial activation to W2 to late responses to W1, delayed by lexical competition. In contrast, the mean lexical-decision latency in the N3W1 condition would include responses triggered by W1 activation, moderately slowed down by the presence of a subcategorical mismatch.

As a first step in evaluating this hypothesis, we tested whether the variability in lexical-decision latencies to W2W1 sequences was greater than the variability in latencies to N3W1 sequences, as our account of the lexical-decision data implies.⁶ We analysed the lexical-decision latencies from McQueen et al. (1999, Experiment 3), using standard deviation as a simple index of variability.⁷ Following McQueen et al., reaction times slower than 1500 ms after sequence offset were excluded, but missing data points, due to outlying RTs or to errors, were not replaced. Latency standard deviations per participant and per condition were then computed and submitted to a one-way ANOVA. Mean standard deviations were 162 ms, 201 ms, and 168 ms for W1W1, W2W1, and N3W1 conditions, respectively, with a significant effect of conditions ($F(2, 88) = 11.85, p < .0001, MSE = 1700.8$). Newman-Keuls tests indicated that the standard deviation was greater in the W2W1 condition than in the N3W1 condition (with $\alpha = .05$).

The higher latency variability in W2W1 than in N3W1, despite equivalent means, suggests that the responses generated in these conditions may come from different underlying patterns of lexical activation. In order to provide further support for our account of the lexical-decision data, we developed a simple model to simulate these data under the assumption that a “yes” response is generated probabilistically when either the target (W1) or the competitor (W2) reaches threshold. (The activation of other words in the lexicon remains low and could have only negligible impact.) We conducted separate simulations using the eye-movement data collected in Experiment 2 and TRACE-activation data. Each simulation was run 1000 times for each of the 15 experimental items, across a range of response thresholds.⁸ These simulations showed that

⁶ It might seem that a stronger test of our hypothesis would be to test whether the distribution of lexical-decision responses in W2W1 differs from the response distribution in N3W1, and in particular, whether it fits a mixture of the two distributions generated by W2 and W1 activation (this mixture need not result in a bimodal distribution, in contrast with what Norris et al. [2000, p. 356] have argued). However, because the W2W1 and N3W1 distributions are very similar when the data from all the subjects are combined, and the number of data points per subject per condition is very small, explicit tests of the hypothesis are difficult.

⁷ We are grateful to James McQueen, Dennis Norris, and Anne Cutler for providing their raw data.

⁸ Space constraints preclude extensive details on the algorithm used in the simulations but a fuller report can be obtained from the authors.

predicted lexical-decision means given W2W1 and N3W1 were roughly identical and higher than for W1W1 across a range of thresholds before diverging, whether computed from fixation proportions from Experiment 2 or from TRACE activation. These simulations demonstrate that identical lexical-decision means can be obtained from different underlying W1 and W2 activation patterns in the W2W1 and N3W1 conditions, if one assumes that “yes” responses can be triggered by the activation of either the target W1 or its close competitor W2.

It is important to note that our lexical-decision simulations do not provide a comprehensive account of the McQueen et al. (1999) and Marslen-Wilson and Warren (1994) data. In particular, the simulation incorrectly predicts a higher error rate for the nonword sequence W2N1 (e.g., *smo(g)b*) than for the sequence N3N1 (e.g., *smo(d)b*) because early activation of W2 would trigger some “yes” responses. The human data showed low and equivalent error rates for both conditions. However, a closer look at the error rates reveals a complex pattern. Error rates were substantially higher in W2W1 and N3W1 than in all the other conditions, in both Marslen-Wilson and Warren (1994) and McQueen et al. (1999). This suggests a bias for the listeners to respond “no” to stimuli containing mismatching coarticulatory cues. Neither our lexical-decision simulation nor Norris et al.’s model accounts for these high error rates. A complete simulation of the lexical-decision data, accounting for both response latencies and error rates, will require a more complex decision mechanism than a simple threshold. The decision rule adopted by participants in these studies may well have been affected by the unusual nature of the stimuli, which contained a large proportion of cross-spliced sequences.

Regardless of how the lexical-decision data are interpreted, simulations of lexical activation of cross-spliced sequences with subcategorical mismatches all assume that, at some point during the recognition process, the recognition of the target W1 is more impaired given W2W1 than N3W1. The lexical-competition effect is apparent both in the TRACE simulations presented here and in Merge’s activation levels in Norris et al. (2000), where word nodes in these localist connectionist models directly compete via lateral inhibition. These models are thus consistent with our eye-movement data. As shown in its simulations of the Marslen-Wilson and Warren data, the distributed model of speech perception (Gaskell & Marslen-Wilson, 1997, Figure 3) also predicts lexical-competition effects during the recognition of the target W1, although they are predicted to be small and transient. Our eye-movement data are also potentially consistent with the Cohort model (Marslen-Wilson, 1987) if one assumes that eye movements do not reflect lexical-activation levels per se, but are generated probabilistically from a *continuously* operating decision mechanism. A decision rule would continuously evaluate a candidate’s activation

proportionally to its competitors' activation. This would predict delayed target recognition in W2W1, because the most active competitor (W2) would be more active in W2W1 than in N3W1 (assuming other competitors were equally active). However, our eye-movement data rule out a model where lexical competition takes place at a late decision stage of recognition, because lexical-competition effects were observed well before the end of the cross-spliced stimuli.

We conclude by highlighting the critical implications of the current work. First, our results provide clear evidence for lexical competition. Second, because most current models incorporate some form of lexical competition, distinguishing among competing models will require detailed information about the time course of lexical activation, including how quickly competition is resolved, and an explicit hypothesis linking behavioural data to underlying processes. Our results demonstrate that the eye-tracking paradigm meets these criteria when coupled with simulations from an explicit model. Finally, the present study provides information about the time course of lexical competition that may provide important constraints on models of spoken word recognition.

REFERENCES

- Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Cutler, A., Norris, D., & McQueen, J. (2000). Tracking TRACE's troubles. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Proceedings of the Workshop on Spoken Word Access Processes* (pp. 63–66). Nijmegen: Max Planck Institute for Psycholinguistics.
- Cycowicz, Y.M., Friedman, D., Rothstein, M., & Snodgrass, J.G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, 65, 171–237.
- Dahan, D., Magnuson, J.S., & Tanenhaus, M.K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cognitive Psychology*, 42, 317–367.
- Dahan, D., Swingle, D., Tanenhaus, M.K., & Magnuson, J.S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, 42, 465–480.
- Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (pp. 31–45). New York: Springer-Verlag.
- Frauenfelder, U.H., & Floccia, C. (1998). The recognition of spoken words. In A.D. Friederici (Ed.), *Language comprehension: a biological perspective* (pp. 1–40). Berlin: Springer.
- Frauenfelder, U.H. & Peeters, G. (1998). Simulating the time course of spoken word recognition: an analysis of lexical competition in TRACE. In J. Grainger and A.M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 101–146). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

- Gaskell, M.G., & Marslen-Wilson, W.D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes*, *12*, 613–656.
- Gaskell, M.G., & Marslen-Wilson, W.D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science*, *23*, 439–462.
- Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing*, *19*, 1–36.
- Luce, R.D. (1959). *Individual choice behavior*. New York: Wiley.
- Magnuson, J.S., Dahan, D., & Tanenhaus, M.K. (2001). On the interpretation of computational models: The case of TRACE. In J.S. Magnuson & K.M. Crosswhite (Eds.), *University of Rochester Working Papers in the Language Sciences*, *2*(1), 71–91.
- Magnuson, J.S., Tanenhaus, M.K., Aslin, R.N., & Dahan, D. (submitted). The time course of spoken word recognition in an artificial lexicon.
- Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.
- Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing. Psycholinguistic and computational perspectives* (pp. 148–172). Hove, UK: Lawrence Erlbaum Associates Ltd.
- Marslen-Wilson, W., & Warren P. (1994). Levels of perceptual representation and process in lexical access. *Psychological Review*, *101*, 653–675.
- McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.
- McQueen, J.M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1363–1389.
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.
- Norris, D., McQueen, J.M., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–325.
- Saslow, M.G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*, 1030–1033.
- Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.
- Streeter, L.A., & Nigro, G.N. (1979). The role of medial consonant transitions in word perception. *Journal of Acoustical Society of America*, *65*, 1533–1541.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.
- Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 353–393). Amsterdam: Elsevier.
- Whalen, D.H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics*, *50*, 351–360.
- Whalen, D.H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, *35*, 49–64.
- Zwitserslood, P., & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, *10*, 121–136.

APPENDIX A

Same voicing

Target W1	Word Competitor W2	Nonword Competitor N3
net (<i>git</i>)	neck (<i>gik</i>)	*nep (<i>gip</i>)
tap (<i>tap</i>)	tack (<i>tak</i>)	tat (<i>tat</i>)
bud (<i>bud</i>)	bug (<i>bug</i>)	*bub (<i>bub</i>)
butt (<i>b^t</i>)	buck (<i>b^k</i>)	*bup (<i>b^p</i>)
carp (<i>kup</i>)	cart (<i>ku</i>)	*cark (<i>kuk</i>)
fort (<i>srt</i>)	fork (<i>srk</i>)	*forp (<i>srp</i>)
harp (<i>lrp</i>)	heart (<i>lrt</i>)	hark (<i>lrk</i>)

Different voicing

Target W1	Word Competitor W2	Nonword Competitor N3
cat (<i>kat</i>)	cab (<i>kab</i>)	*cag (<i>kag</i>)
bat (<i>bat</i>)	bag (<i>bag</i>)	*bab (<i>bab</i>)
road (<i>rid</i>)	rope (<i>rip</i>)	*roke (<i>rik</i>)
pit (<i>pit</i>)	pig (<i>pig</i>)	*piib (<i>pib</i>)
hood (<i>sud</i>)	hook (<i>suk</i>)	*hoop (<i>sup</i>)
knot (<i>gut</i>)	knob (<i>gub</i>)	*knog (<i>gug</i>)
beak (<i>bik</i>)	bead (<i>bid</i>)	*beab (<i>bib</i>)
rod (<i>rud</i>)	rock (<i>ruk</i>)	*rop (<i>rup</i>)

Note: * indicates a sequence that is not a real word in American English. The transcriptions adopted in the TRACE simulations (given the lack of some phonemes) are indicated in parentheses.

APPENDIX B

Experimental Trials

<i>Distractor 1</i>	<i>Target W1</i>	<i>Distractor 2 / Competitor W2</i>	<i>Distractor 3</i>
bass	net	nurse / neck	deer
skunk	tap	trunk / tack	peas
fox	bud	bow / bug	eye
clams	butt	bride / buck	ghost
swing	carp	comb / cart	moon
light	fort	flag / fork	hat
desk	harp	house / heart	claw
vase	cat	crown / cab	tree
pen	bat	bone / bag	stool
knee	road	rug / rope	glass
ark	pit	pot / pig	flute
eggs	hood	hose / hook	brush
mouse	knot	knight / knob	beer
saw	beak	bench / beads	thumb
bear	rod	rake / rocks	fries

Note: Distractor 1 corresponds to the target picture from the initial “point to” instruction. Distractor 2 was displayed only in Experiment 1; the Competitor W2, only in Experiment 2.