

TRACKING THE TIME COURSE OF SUBCATEGORICAL MISMATCHES ON LEXICAL ACCESS: EVIDENCE FOR LEXICAL COMPETITION

D. Dahan

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

J. S. Magnuson, M. K. Tanenhaus, and E. M. Hogan

University of Rochester, Rochester, New York

ABSTRACT

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"). Latency to fixate the target picture was slower when the stimulus onset originated from a competitor word (e.g., ne(ck)t) than from a nonword (e.g., ne(p)t), reflecting lexical competition. Furthermore, simulations with the TRACE model of spoken-word recognition captured the major trends of fixation probabilities to the target and its competitor. We argue that the fixation functions provide a fine-grained measure of underlying activations and can reveal lexical-competition effect that other paradigms, such as the lexical-decision task, conceal.

1. INTRODUCTION

It is now generally accepted that as listeners attend to a spoken word, they are simultaneously entertaining several candidates with phonological representations that match the spoken input, and these candidates compete for recognition. The degree of activation of each candidate varies with the goodness of fit between the input and its representation. How the competition between active candidates is realized and resolved is subject to much debate. Models like TRACE [1] or Shortlist [2] assume that candidates that match the same portion of the input actively compete with each other via a lateral-inhibition mechanism. On the other hand, the Cohort model [3] does not assume any lateral inhibition, so that the presence of a competitor does not directly affect the activation level of a word candidate. A word is recognized when its activation level becomes critically different from the activation level of its competitors. Competition is thus assumed to take place at a decision stage.

Marslen-Wilson and Warren [4] provided evidence that they interpreted as inconsistent with competition operating via lateral inhibition. They created cross-spliced word sequences whose initial CV portion had been excised from another token of the same word (e.g., jo(b)b), from another existing word (e.g., jo(g)b), or from a nonword (e.g., jo(d)b). These word sequences were auditorily presented to participants who performed a lexical decision. The lexical-decision mean latency to the jo(g)b sequence did not differ from that to the jo(d)b sequence, and both were significantly slower than the latency to the jo(b)b sequence. The authors interpreted this result as strong evidence against models that allow

lexical competition via lateral inhibition. They reported simulated lexical activations in TRACE that, they argued, were unable to account for the human data. The response probability for the target *job* was lower in jo(g)b than in jo(d)b because the initially activated competitor *jog* inhibited the target *job*, while this inhibition was much reduced when the pre-splice section came from the nonword *jod*, which weakly supports both *jog* and *j.b*. In TRACE, inhibition *modifies* the activations of words throughout processing. By contrast, in a model like Cohort, a word activation is determined only by its goodness of fit to the input. However, via competition at the decision stage, the degree of activation of the competitor *jog* could in principle influence the recognition of *job*, unless the activation of *job* had surpassed that of *jog* by the time the lexical decision was performed.

Thus, the absence of difference between the jo(g)b and jo(d)b conditions cannot be taken as evidence against lateral inhibition. Indeed, McQueen, Norris, and Cutler [5], who replicated this empirical result, argued that it can fall out of a model that allows lateral inhibition between active lexical candidates. Norris, McQueen, and Cutler [6] simulated the activations of *job* and its competitor *jog* in a two word-node network in the three splicing conditions. In the jo(b)b condition (thereafter W1W1 condition), the target *job* quickly reached its peak, with little competition from *jog*. In the jo(d)b condition (N3W1), both *job* and *jog* were initially equally active; when the information supporting the last consonant "b" was provided, *job* surpassed *jog* and reached the response threshold of 0.2 at time slice 9.4. In the jo(g)b condition (W2W1), *jog* was first highly activated and strongly inhibited *job*; when the information supporting the last consonant was provided, the tendency was inverted and the activation of *job* sharply increased, reaching the response threshold at time slice 9.7. The similarity between the time slices at which the W2W1 and N3W1 conditions reached threshold supported the lexical-decision data. Norris et al. argued that, because the lexical competition between *jog* and *job* is very quickly resolved, the activation of the target *job* reaches the threshold that triggers lexical-decision responses with the same delay as when no such competition takes place.¹ However, to simulate the

¹ The resolution of the competition is partly due to the fact that the word level is allowed to cycle through 15 iterations on each time slice, followed by a reset of lexical-activation levels before the next slice is processed (see [6] for further details)

lexical-decision data, their model depends on choosing a threshold within an extremely restricted range. Interestingly, the threshold of 0.2 appears to be the lowest value that is not reached by the early activation of the competitor *jog* in the W2W1 condition. Adopting a lower threshold would lead to very early responses triggered by the activation of *jog*.

In fact, we argue that mean lexical-decision latencies may not be an appropriate measure for the activation of a target item, because the lexical-decision task does not require correct identification of the intended target. Subjects may also respond 'yes' in response to the high activation of a competitor item. It may thus be problematic to relate lexical decisions to the underlying activation functions. By contrast, tracking eye movements can provide a continuous measure of lexical activation over time. In this paradigm, participants' eye movements to pictured objects are recorded as they follow instructions to click on one of the four objects (e.g., "click on the net"). Allopenna, Magnuson, and Tanenhaus [7] demonstrated that the proportion of fixations to each picture as the target word is heard and processed can be mapped onto lexical-activation functions using a simple linking hypothesis. The goal of the present study was two-fold. First, we aimed to demonstrate that the eye-tracking paradigm, by contrast with the lexical-decision task, captures the lexical competition taking place in the W2W1 condition. Second, we aimed to confirm the paradigm's linking hypothesis by showing that fixation functions mirror the underlying activation functions generated by TRACE.

2. EXPERIMENT 1

The goal of this experiment was to demonstrate that eye movements, by contrast with lexical-decision latencies, can capture lexical competition in the splicing condition W2W1. To do so, we monitored participants' eye movements to pictured objects as they heard the referent's name (e.g., "net") in each of the three splicing conditions. We hypothesized that the latency with which participants performed an eye movement to the target picture would reflect the target's lexical activation.

2.1. Method

We selected 15 triplets composed of two real words and a nonword ending with a stop consonant. The two words were picturable nouns. In the triplet, one real word was assigned the role of target, and the other, the role of competitor. The voicing feature of the stop consonant was kept constant for all three items for 7 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 8 triplets (e.g., pit, pig, *pib). Each item of the triplets was recorded, digitized, and cross-spliced to yield three splicing conditions for each target item (e.g., ne(t)t, ne(ck)t, ne(p)t). The mean duration of the pre-splice fragment was 379 ms, the duration of the final consonant, 206 ms.

Participants were 30 native speakers of English. They were facing a computer display composed of four line-drawings: the target picture (e.g., a net) and 3 distractor pictures (e.g., nurse, bass, and deer). Note that the picture of the competitor W2 (e.g., neck) was not present. Participants were auditorily instructed to point to one of the distractor pictures with the mouse cursor (e.g., 'point to the bass'). As soon as the cursor reached the picture, the critical instruction was played (e.g., 'now the net'). This procedure aimed to maximize the proportion of trials where participants were still fixating the distractor picture when the critical instruction was heard. We measured the latency with which participants performed an eye movement to the target picture before pointing to it in each splicing condition. In addition, the proportions (across subjects) of fixations to the target picture over time were collected.

2.2. Results and Discussion

For 7 participants, one trial was missing because of technical failure. In the latency analysis, we excluded the few trials where participants were already fixating the target picture at target onset (21 out of 443 trials, 4.7%). On average, the latency to fixate the target picture was 638 ms in the W1W1 condition, 851 ms in the W2W1 condition, and 673 ms in the N3W1 condition. A two-way ANOVA (splicing condition \times voicing status within the triplet) revealed a significant effect of splicing condition ($F_1(2,58) = 13.7, p < .001, MSE = 57092.5$; $F_2(2,26) = 12.5, p < .001, MSE = 14154.3$), but no effect of voicing and no interaction. Newman-Keuls tests indicated that the latency was significantly slower in the W2W1 condition than in the W1W1 and N3W1 conditions, with no significant difference between the W1W1 and N3W1 conditions (with $\alpha = .05$).

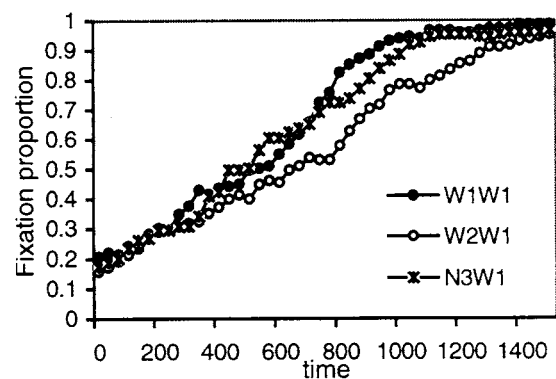


Figure 1. Fixation proportion to the target picture W1 over time since target onset (in ms) for each splicing condition (W1W1, W2W1, N3W1).

The proportions of fixations to the target picture over time for each splicing condition are presented in Figure 1. As can be seen in the figure, fixations between conditions are comparable until about 600 ms after target onset, where the fixations in the W2W1 condition start diverging from those in the W1W1 and N3W1 conditions. Recall that the duration of the pre-splice

fragment was about 400 ms, with coarticulatory cues being strongest presumably in the late portion of the vowel. Given a 150 to 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 W1 were considerably delayed; when this information did not match a word, as in N3W1, no such delay was observed. Fixations to the target in the N3W1 condition were only slightly delayed relative to those in the W1W1 condition.

Experiment 1 suggests that eye movements to the target picture can capture lexical competition in the mismatching condition W2W1. Early in the W2W1 sequence, the competitor W2 becomes highly active and competes with W1. Evidence for the competitor's temporary activation had been provided by gated presentations of W2W1 ([4], [5]), but only eye-movement data revealed its influence on-line.

Fixation latencies in the N3W1 condition did not differ significantly from those in the W1W1 condition, although they tended to be slower. The time-course analysis also showed a small difference between the W1W1 and N3W1 functions. This suggests that lexical access is much more disrupted when mismatching coarticulatory information matches a word than when it matches a nonword.

3. EXPERIMENT 2

The goal of Experiment 2 was to evaluate the fit between the fixation functions obtained in the eye-tracking paradigm and the underlying activations to the target W1 and its cohort competitor W2 generated by a model incorporating lateral inhibition. In order to obtain time-course data on the activation of both the target and the cohort in each of the splicing conditions, we presented the cohort picture along with the picture of the target and two distractors. The target and cohort activations for each splicing condition were simulated using TRACE.

3.1. Eye-tracking study

3.1.1. Method

The materials and procedure used were identical to those used in Experiment 1. The only difference consisted in presenting the picture of the cohort W2 along with the target picture and two distractor pictures (e.g., the pictures of a net, a neck, a bass, and a deer).

Participants were 30 native speakers of English, who had not taken part into Experiment 1.

3.1.1. Results

Because of technical failure, 14 trials (out of 450) were missing. Figure 2 presents the fixations to the target W1 and its competitor W2 for each splicing condition. Fixations to the target over time indicated a fast rise in the W1W1 condition, intermediate in N3W1, and slowest in W2W1. Fixations to the cohort competitor W2 revealed a complementary picture. The competitor

picture was fixated most in the W2W1 condition, 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 favors W1. This pattern of results indicates that fine-grained information in the input is reflected in observed fixations.

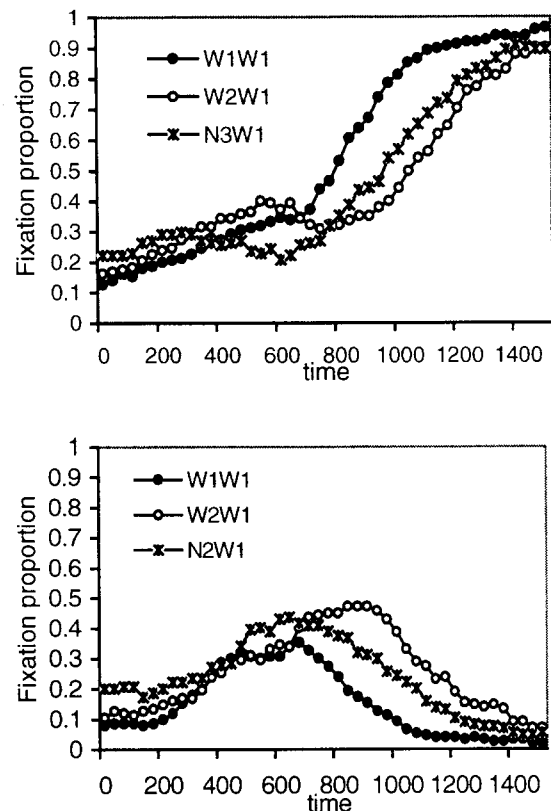


Figure 2. Fixation proportion to the target picture W1 (top panel) and to the cohort picture W2 (bottom panel) over time since target onset (in ms) for each splicing condition (W1W1, W2W1, N3W1).

3.2. Simulations

We used the publicly available TRACE implementation (<ftp://www.crl.ucsd.edu/pub/nnets>) with the standard parameter set reported in [1], and with the lexicon augmented to include transcriptions of our stimuli (for a total of 257 words in the lexicon). TRACE simulates coarticulation by allowing features from one segment to spread over many cycles, possibly beyond the center of the following segment. It is thus not clear how one should define vowel offset from the W1, W2, and N3 inputs. Our stimuli were cross-spliced at the *latest* time-slice during vowel presentation that still provided a basis for clear recognition of W1 after complete input presentation (slice 24, one slice before the center of the next segment). Activation functions generated from the cross-spliced sequences were then converted into predicted fixation probabilities over time using the Luce choice rule applied to the four visually present alternatives (i.e., W1, W2, and two distractors; see [8])

for further details). Figure 3 presents predicted fixation probabilities over cycles.

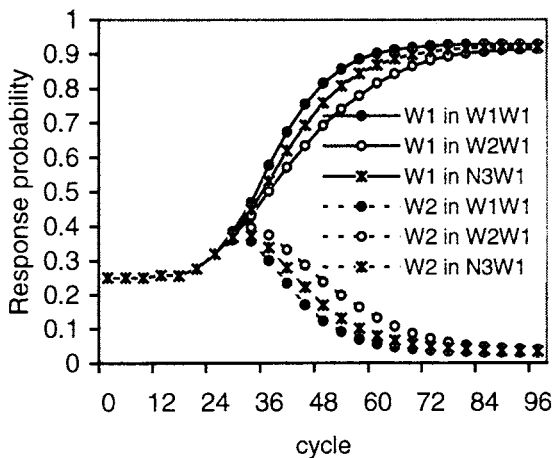


Figure 3. Response probabilities over cycles for the target W1 and the cohort competitor W2, in each of the splicing conditions (W1W1, W2W1, N3W1).

As is apparent from the figure, the probability of fixating the target W1 rises fastest given W1W1, intermediate given N3W1, and slowest given W2W1; fixations to the cohort competitor W2 indicates more fixations given W2W1, less given N3W1, and the least given W1W1. The predicted fixations mirror the human data quite closely, although the magnitude of the effect is reduced compared to the human data, under TRACE's default parameter set. To our surprise, we were unable to replicate the simulations reported by Marslen-Wilson and Warren ([4], Figures 12 and 13), where the probability of recognizing W1 was very low given W2W1 and comparable given N3W1 and W1W1. We suspect that the discrepancy originates from the cross-splicing point chosen. If this point is too far into the following segment, the system receives too much information supporting W2, and is unable to produce the appropriate recognition behavior. For example, W2W1 would be recognized as W2 rather than W1.

A possible aspect of TRACE that can account for the difference in magnitude between the human data and the simulations is the resolution of the input. By default, there are only six input slices between the peaks of adjacent segments. This gives a very small window for subcategorical mismatches to take effect. In initial explorations, we have found that increasing feature spreading and/or phoneme durations improves the fit with our data.

4. DISCUSSION

The results from both experiments and the simulations suggest that eye-movement data provide a fine-grained measure of underlying activations and can reveal lexical competition. Fixations to the target picture were more delayed when the target word was cross-spliced from an existing word than from a nonword. This result contrasts with lexical-decision data, which showed equivalent

performances in both cross-splicing conditions. The lexical-decision data have been interpreted as evidence against lexical competition via lateral inhibition ([4]). However, as discussed in our introduction, this interpretation is untenable. The lexical-decision data have also been interpreted as reflecting the resolution of lexical competition before a response is generated ([6]). The present data and simulations suggest that lexical competition takes place and that its effects persist for a substantial amount of time beyond the offset of a stimulus. We propose that the lexical-decision task conceals the lexical-competition effects because the activations of competitors as well as targets influence responses. On the basis of the activation functions, we are currently developing a model that simulates the lexical-decision data across a range of response thresholds, under the assumption that a 'yes' response is triggered probabilistically given high activation of any item. This model makes fine-grained predictions, such as similar mean lexical decision latencies given W2W1 and N3W1, but greater variability given W2W1 due to early responses based on the activation of W2. Such predictions are needed to link activation-based models to task-specific decision models.

5. ACKNOWLEDGEMENTS

This work was supported by NSF grant SBR-9729095 to M.K.T. and an NSF GRF to J.S.M. Corresponding author: Delphine Dahan, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands. Email: delphine.dahan@mpi.nl.

6. REFERENCES

- [1] McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- [2] Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.
- [3] Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, **25**, 71-102.
- [4] Marslen-Wilson, W. & Warren P. (1994). Levels of perceptual representation and process in lexical access. *Psychological Review*, **101**, 653-675.
- [5] 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.
- [6] Norris, D., McQueen, J. M. & Cutler, A. (in press). Merging information in speech recognition: feedback is never necessary. *Behavioral & Brain Sciences*, **23**.
- [7] 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.
- [8] Dahan, D., Magnuson, J. S. & Tanenhaus, M. K. (submitted). Time course of frequency effects in spoken-word recognition: evidence from eye movements.