Tracking the Time Course of Lexical Activation in Continuous Speech

Paul D. Allopenna (ALLOPEN@BCS.ROCHESTER.EDU)
James S. Magnuson (MAGNUSON@BCS.ROCHESTER.EDU)
Michael K. Tanenhaus (MTAN@BCS.ROCHESTER.EDU)
Department of Brain and Cognitive Sciences
University of Rochester, Meliora Hall, Rochester, NY 14627 USA

Abstract

Eye-movements to pictures of four objects on a screen were monitored as participants heard progressively larger gates and tried to identify the object (Experiment 1) or followed a spoken instruction to move one of the objects (Experiment 2), e.g., "Pick up the beaker; now put it below the diamond". The distractor objects included a cohort competitor with a name that shared the initial onset and vowel as the target object (e.g., *beetle*), a rhyme competitor (e.g. *speaker*) and an unrelated competitor (e.g. *carriage*). In the gating task, which emphasizes word initial information, there was clear evidence for multiple activation of cohort members, as measured by judgments and eye-movements. With continuous speech there was clear evidence for both cohort and rhyme activation as predicted by continuous activation models such as TRACE (Elman and McClelland, 1988). Moreover, the time course and probabilities of eye-movements closely corresponded to simulations generated from TRACE.

Introduction

Recognition of a spoken word often occurs prior to the end of the word and is not only influenced by the properties of the word itself (e.g., its frequency), but also by the set of words to which it is phonetically similar (e.g., its lexical neighborhood). The cohort model, developed by Marslen-Wilson and colleagues (e.g., Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987), accounted for these phenomena by proposing that the onset of a word activates a set of lexical candidates which compete for recognition. Thus, as the word *beaker* is spoken, both *beaker* and *beetle* will initially become active members of a recognition cohort. Members of the cohort are subsequently evaluated using contextual information and subsequent speech input. Extensive empirical evidence now supports the claim that words with shared onsets are briefly activated during spoken word recognition. For example, lexical decisions to visually presented associates of cohort members are facilitated when targets are presented early in a word (e.g., Marslen-Wilson & Zwitserlood, 1989; Zwitserlood, 1989). Thus, *beaker* would initially prime both *glass* (associate of *beaker*) and *bug* (associate of *beetle*).

However, the cohort model makes some problematic assumptions. First, word onsets are often not clearly marked in the speech stream, especially in continuous speech. Second, lexical candidates that have only a partial match to the onset of the word will never enter into the recognition set. This will limit the robustness of the model in noisy environments

Continuous recognition models, such as the TRACE model (McClelland & Elman, 1986), the Shortlist model (Norris, 1994) and the Neighborhood Similarity Model (Goldinger, Luce, & Pisoni, 1986), overcome these problems by assuming that (a) lexical access takes place continuously and (b) recognition is based on the similarity of the unfolding speech to lexical representations. The initial portion of a spoken word will still exert a strong influence on alternatives that are activated shortly after the word begins. However, the set of activated alternatives will also include words that do not share initial onsets. This has the desirable property of allowing lexical access to be successful without assuming that onsets are clearly marked in the speech. It also leads to a more error-tolerant system, because lexical representations that are not initially activated can still accrue activation if their overall similarity to the input is high.

Unfortunately, evidence from on-line tasks for activation of lexical candidates which do not share onsets is quite weak (cf. Zwitserlood, in press). For example, the evidence for activation of associates of words that rhyme with a spoken prime is equivocal (e.g., *beaker* priming an associate of *speaker*). Finding evidence for lexical competitors is further complicated by paradigm limitations. Semantically-mediated cross-modal priming is a relatively indirect way of assessing lexical activation; a candidate may not prime an associate until it becomes highly active, limiting the sensitivity of the measure. In addition, the mechanism linking cross-modal priming to underlying word recognition processes remains underspecified (cf. McKoon, Allbritton & Ratcliff, 1996). Other results that support the predictions of continuous activation models are problematic in that they derive from paradigms that either involve meta-linguistic judgments or do not allow recognition to be monitored as the speech unfolds.

We have been exploring a paradigm in which participants follow spoken instructions to touch or manipulate objects in a visual workspace while we monitor eye-movements using a lightweight camera mounted on a headband (Tanenhaus et al, 1995). Saccadic eye-movements are extremely frequent, and thus have the potential of providing a sensitive measure of recognition processes. Moreover, monitoring eye-movements allows one to observe lexical access during continuous speech without requiring subjects to make an overt decision.

In previous work, we established that the paradigm is sensitive to cohort effects. Spivey-Knowlton and Tanenhaus (Tanenhaus et al., 1995) had participants follow instructions to pick up and move objects on a horizontal board. The set of objects sometimes included an object with a name beginning with the same phonetic sequence as the target object. Examples of objects with overlapping initial phonemes were *candy* and *candle*, and *cart* and *carton*. The presence of a competitor increased the latency of eye-movements to the target and induced frequent looks to the competitor. The timing of these eye-movements indicated that they were programmed during the "ambiguous" segment of the target word. These results demonstrated that the two objects with similar names were, in fact, competing as the target word unfolded.

The current research extended these investigations in three important ways:

(1) We found evidence against the hypothesis that using a circumscribed visual world would artificially increase activation to alternatives that should not be normally activated. (2) We demonstrated that lexical alternatives that do not share onsets (e.g., rhymes) are partially activated during continuous speech. (3) We used simulations from a computationally explicit continuous activation model, TRACE, to show that the probability of an eye-movement being generated to a target object is directly related to its activation level.

Experiment 1 used a "gating" paradigm in which participants heard successively longer fragments of a word on each gate. Their task was to point to which of the four pictures was being named. On critical trials, the pictures included the target (e.g., *beaker*), a cohort competitor with a shared onset (e.g., *beetle*), a rhyme competitor (e.g., *speaker*), and an unrelated item (e.g., *parrot*). In gating, only cohort members are typically generated as responses. This is not surprising from the perspective of continuous activation models because gating places clear emphasis on the beginnings of words. Evidence for more rhyme choices with gating compared to the unrelated baseline would have suggested that the presence of a limited set was inflating similarity effects. However, no evidence for rhyme choices was found. In addition, the pattern of eye-movements provided strong evidence that multiple alternatives -- the cohort competitor and the target were both activated during the initial gates.

Experiment 2 used the same stimuli with continuous speech instructions (e.g., "Pick up the beaker. Now put it below the diamond."). Probability functions for eye-movements to the target, cohort competitor, rhyme competitor, and the unrelated item were generated by transforming activation levels from TRACE into probability functions. The data from the experiment closely matched the predictions from the simulations.

Experiment 1

Method

Participants. Six male and female students at the University of Rochester were paid for their participation. All were native speakers of English with normal or corrected-to-normal vision.

Materials The stimuli were based on eight "referent-cohort-rhyme" triples. For example, one triple consisted of *beaker* (the "referent" word), its left-to-right cohort, *beetle*, and a rhyme, *speaker*. For each word, there was a corresponding black and white line drawing. Frequency was not controlled for, due to the limited number of words that met our criteria. However, post-hoc analyses in both experiments revealed no frequency effects.

For the experiment itself, stimuli were presented in groups of four on a computer background screen which was divided into a 5 x 5 grid. The center square of the grid contained a cross which the participant was asked to fixate on until the presentation of the auditory stimulus. The line drawings for each trial were placed in the squares on the grid that were diagonally adjacent to the center cross.

There were 16 total words presented in the experiment. Eight of the words were presented in "critical" trials (i.e., trials with both cohort and a rhyme competitors, as well as an unrelated item), and eight of the words were presented in "filler" trials. For both critical and filler presentations, there were between 8 and 10 gates per word. The first gate of each word varied in time depending on the initial segments of the word. The first gate started at the word onset and ended at the fourth zero-crossing after vowel onset. Thereafter, each gate added an additional 40 ms onto the preceding gate.

Procedure Participants were seated at a comfortable distance from a computer. Prior to the experiment, participants were twice shown pictures of the stimuli they were to see in the experiment. First they were shown each stimulus picture with its name written underneath the picture. Subsequently, they were shown each stimulus picture, but this time without its name. In both cases, the stimuli were presented in random order. During the second viewing, participants were asked to name each of the objects aloud. Subjects were corrected if they mistakenly named an object. With one exception, participants correctly named all of the stimuli on their first attempt.

Prior to hearing the first gate for any given word, participants were shown the grid containing drawings of the four objects relevant to that trial. As soon as they were ready to proceed, they signaled to the experimenter, who then began the first gating presentation. For each gate the stimulus screen was first displayed for approximately 2 seconds. During this time the experimenter instructed the participants to fixate the center cross. Once the auditory stimulus was presented, participants indicated which word they thought they heard by touching the object on the computer screen whose named matched their hypothesis.

We tracked eye-movements with an Applied Scientific Laboratories (E4000) eye tracker. Two cameras mounted on

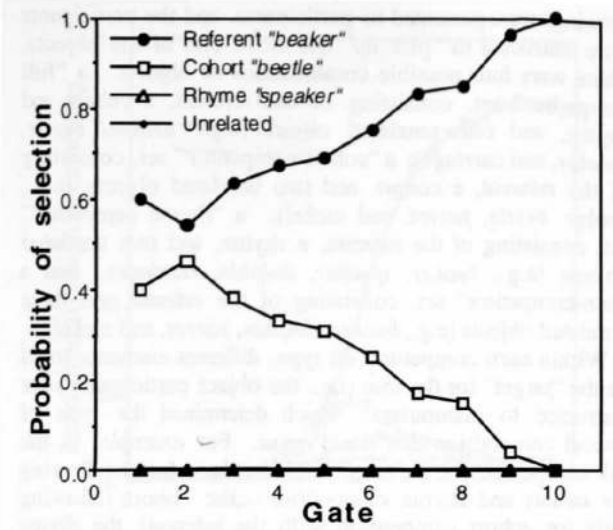

Figure 1: Probability of selecting each item at each gate in Experiment 1.

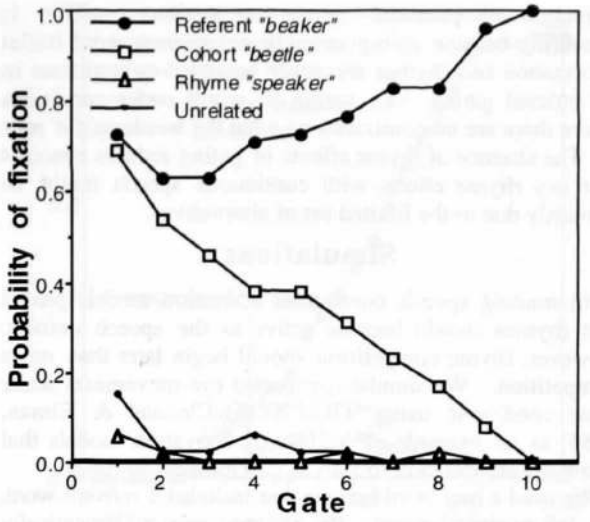

Figure 2: Probability of fixating each item at each gate in Experiment 1.

a lightweight helmet provide the input to the tracker. The eye camera provides an infrared image of the eye. The center of the pupil and the first Purkinje corneal reflection are tracked to determine the orbit of the eye relative to the head. Accuracy is better than 1 degree of arc, with virtually unrestricted head and body movements. A scene camera is aligned with the participant's line of sight.

A calibration procedure allows software running on a PC to superimpose crosshairs showing the point of gaze on a HI-8 video tape record of the scene camera. The scene camera samples at a rate of 30 frames per second, and each frame is stamped with a time code. The auditory stimuli were presented binaurally through headphones using the standard digital-to-analog devices provided with the experimental control computer (an Apple Power Macintosh 7200), as well as through the internal speaker of the computer. A microphone connected to the HI-8 VCR provided an audio record of each trial.

Results and Discussion

Figure 1 shows the probability with which participants selected each of the four pictures for the first eight gates. On the initial gates, the referent and the cohort competitor were equally likely, with the probability of selecting the target increasing across gates. Rhyme and unrelated objects were rarely selected and did not differ. There was a main effect of Response-Type ($F(3, 15) = 621.66$, $p = .0001$), as well as an interaction between Response Type and Gate ($F(9, 45) = 3.05$, $p = .0062$). Individual means comparisons indicated that the Cohort and the Target responses differed beginning at gate 3. These data closely match typical gating data, in that rhymes are almost never generated as lexical candidates.

The eye-movement data confirmed that during the early gates, both the cohort competitor and the referent were being considered. Figure 2 shows the probability of making an eye-movement to each of the pictures across gates. There was a significant main effect of Gate ($F(3, 15) = 4.55$, $p = .019$), a significant main effect of Response-Type ($F(3, 15) = 351.27$, $p = 0.0001$), and a significant interaction between Gate and Response-Type ($F(9, 45) = 2.97$, $p < .0073$). Comparisons of individual means showed no significant differences between Cohort and Target responses until gate four.

Figure 3 shows the probability of making an eye-movement to the cohort, rhyme and unrelated pictures, when the referent was chosen. The high probability of looks to the cohort object confirms that it was being considered even when the referent was chosen. There was no suggestion from the eye-movements that the rhyme was more highly activated than the unrelated word.

The results provide clear evidence for activation of lexical candidates sharing onsets. No evidence was found for

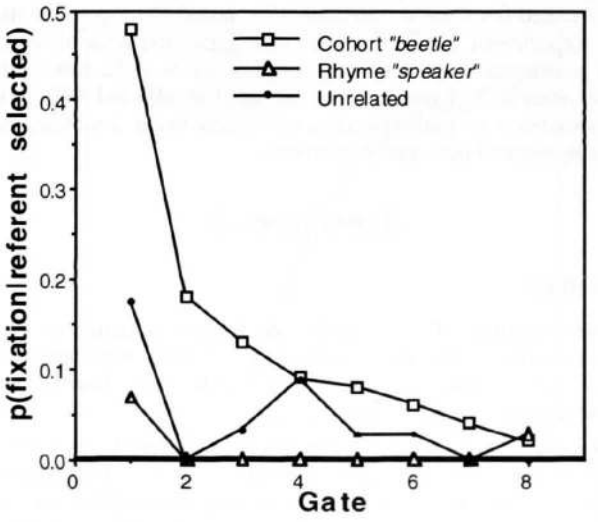

Figure 3: Conditional probability of fixating each item at each gate, given that the referent has been selected.

activation of potential rhyme competitors. This is reassuring because gating strongly emphasizes word initial information and rhymes are rarely generated as responses in unrestricted gating, i.e., gating to words under conditions where there are no constraints on what the words might refer to. The absence of rhyme effects in gating reduces concerns that any rhyme effects with continuous speech might be primarily due to the limited set of alternatives.

Simulations

With running speech, continuous activation models predict that rhymes should become active as the speech unfolds. However, rhyme competition should begin later than onset competition. We simulated expected eye-movements under these conditions using TRACE (McClelland & Elman, 1986) as an example of a class of activation models that allow for competition to start at any moment in time.

We used a four word lexicon that included a referent word, its left-to-right cohort, its rhyme, and a phonetically unrelated word. The referent word input was manipulated such that the first two phonemes were given up to a maximum of 15% noise to approximate a normal speech environment. Input words were run for 72 cycles of processing, with word activations noted every 3 cycles for all four words. Activations were then translated into response strengths, following Luce (1959):

$$S_i = e^{ka_i}$$

where k is a free parameter that determines the amount of separation between units of different activations. Response strengths are then converted into probabilities with the following formula:

$$p(R_i) = S_i / \Sigma S$$

We made the simple linking assumption that the probability of making an eye-movement to an object during on-line processing would be a direct result of its activation. Figure 4 shows the predicted pattern of eye-movements as generated from the simulations. We tested these predictions in Experiment 2. Participants were presented with pictures of four items (e.g., a referent word, a left-to-right cohort of the referent, a rhyme, and an unrelated word), and followed instructions to move pictures of objects using a mouse, as we monitored their eye-movements.

Experiment 2

Method

Participants Twelve male and female students at the University of Rochester were paid for their participation. All were native speakers of English with normal or corrected-to-normal vision.

Materials The stimuli were based on the eight "referent-cohort-rhyme" triples used in Experiment 1. The triples were divided into four pairs, which were presented to different groups of participants. On any trial, four line drawings were presented to participants, and the participants were instructed to "pick up" and move one of the objects. There were four possible combinations of objects: a "full competitor" set, consisting of the referent, a cohort and rhyme, and one unrelated object (e.g., *beaker*, *beetle*, *speaker*, and *carriage*); a "cohort competitor" set, consisting of the referent, a cohort, and two unrelated objects (e.g., *beaker*, *beetle*, *parrot*, and *nickel*); a "rhyme competitor" set, consisting of the referent, a rhyme, and two unrelated objects (e.g., *beaker*, *speaker*, *dolphin*, *carriage*); and a "non-competitor" set, consisting of the referent and three unrelated objects (e.g., *beaker*, *dolphin*, *parrot*, and *nickel*).

Within each competitor set type, different elements could be the "target" for the trial (i.e., the object participants were instructed to manipulate), which determined the type of lexical competition that could occur. For example, in the full competitor set, the target could be the referent (allowing for cohort and rhyme competition), the cohort (allowing only for cohort competition with the referent), the rhyme (allowing only for rhyme competition with the referent), or the unrelated object (which should eliminate competition).

Procedure Participants were seated at a comfortable distance from a computer. Before the first trial, participants were shown pictures of the eight stimuli they were to see in the experiment. The experimenter named them once, and then asked the participant to name them. This was repeated until the participant named every object correctly.

The structure of each trial was as follows. First, a five-by-five grid appeared. Then, line drawings of the stimuli appeared on the grid. After approximately 1 second, the experimenter would instruct the participant to look at the center cross. Participants were instructed before the experiment began that they could move their eyes freely up until this instruction, but then were to fixate the cross until the next instruction. After approximately one more second, the experimenter instructed the participant to pick up one of the objects (e.g., "pick up the beaker"). Once the participant had "picked up" the object (by clicking on it once using the computer's mouse), the experimenter instructed the participant to place it "next to", "above", or "below" one of four geometrical figures which appeared in fixed locations on every trial (e.g., "now put it above the triangle"). Once the participant had placed the object in the appropriate square, the experimenter instructed the participant to look at the center cross. When the participant indicated that she was fixating the cross by clicking on it with the mouse, the trial ended. The grid was then replaced by a blank white screen followed by the calibration screen. Participants could then request to take a break, or the equipment calibration could be verified.

Results and Discussion

The results of Experiment 2 indicate that both cohorts and rhymes were active candidates for recognition, as predicted by TRACE. Figures 5-7 plot the probability over time of fixations on particular objects.

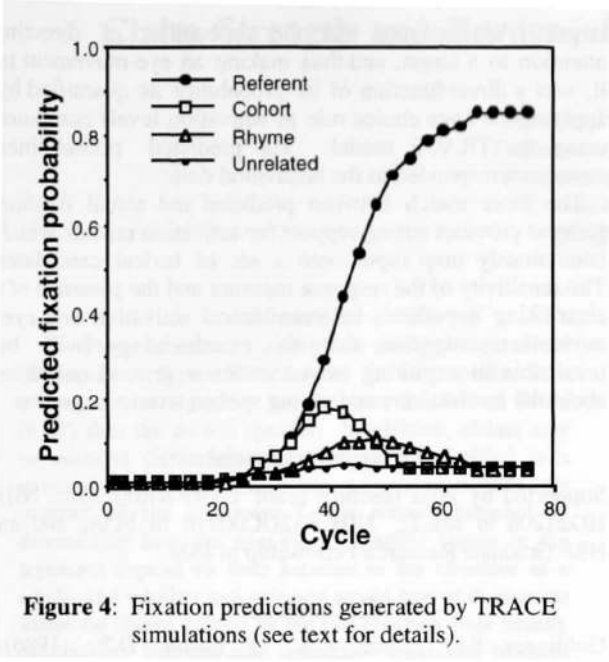

Figure 4: Fixation predictions generated by TRACE simulations (see text for details).

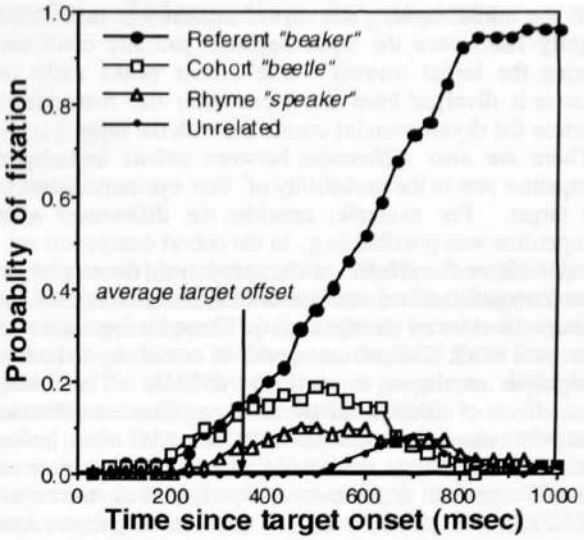

Figure 5: Probability of fixating each item in Experiment 2 when the referent is the target, and both a cohort and a rhyme are present.

As can be seen in Figure 5, when both a cohort and a rhyme were present, participants were as likely to look at a cohort as they were the target item until approximately 430 ms after the onset of the target word. They were also much more likely to fixate a rhyme than an unrelated item, although the separation of rhyme fixation probability from the unrelated baseline was somewhat delayed relative to the referent and cohort (again, as predicted). Given that at least 150 ms are required to plan an eye-movement, and that the average target word duration was 335 ms (with a range of 233 to 467 ms), participants seemed to incrementally update hypotheses as to target word identity over time. Participants were highly unlikely to look at an unrelated object (until the rise that occurs after 500 ms, which appears to have been random scanning).

In Figure 6, fixation probabilities over time are plotted when the stimulus set consisted of a referent, a cohort, and two unrelated objects, and in Figure 7, fixation probabilities over time are plotted when the stimulus set consisted of a referent, a rhyme, and two unrelated objects. As can be seen in Figures 5-7, while subjects were more likely to fixate either a cohort or rhyme than they are to fixate unrelated objects, there are some differences between cohorts and rhymes. Cohort activation tends to rise more quickly and have a higher peak, but rhyme activation tends to persist for a longer time. These are the general trends predicted by the TRACE simulations. The referent and cohort probabilities separate from baseline together since both are consistent

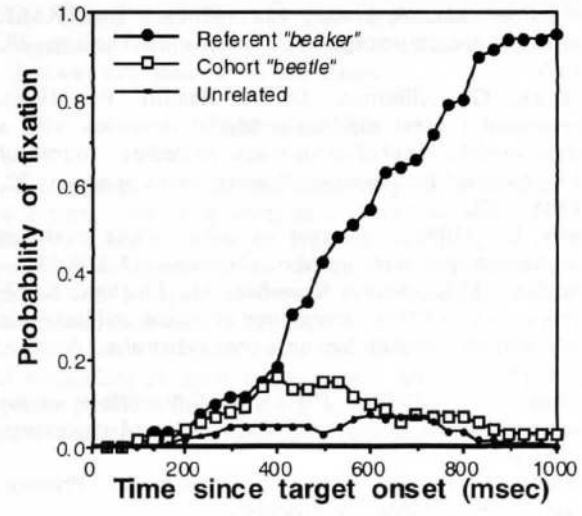

Figure 6: Probability of fixating each item in Experiment 2 when the referent is the target, and a cohort is present.

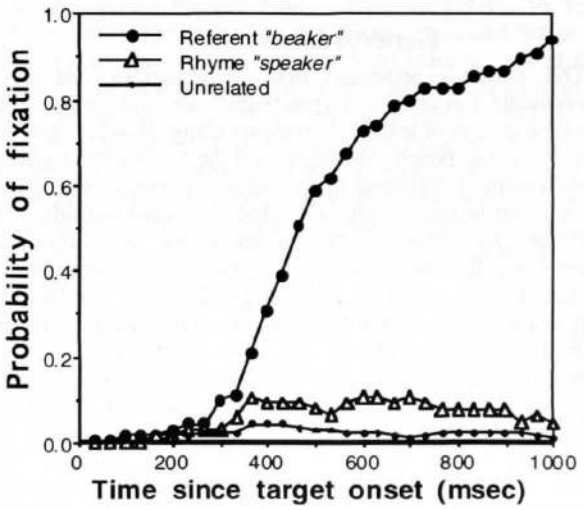

Figure 7: Probability of fixating each item in Experiment 2 when the referent is the target, and a rhyme is present.

with the initial input. The rhyme separates from baseline slightly later, once the input becomes partially consistent (during the initial vowel). The cohort peaks early on because it diverges from the input after the first vowel, whereas the rhyme remains consistent with the input.

There are also differences between cohort and rhyme competitor sets in the probability of first eye-movements to the target. For example, consider the differences when competition was possible (e.g., in the cohort competitor set, because either the referent or the cohort was the target) vs. when competition was not possible (i.e., when one of the two unrelated items was the target). These hit rate data were submitted to a 2 (competition possible or not) by 2 (cohort competitor or rhyme competitor) ANOVA. There were main effects of competition (across competitor sets, hit rate was .649 competition was possible, vs. .855 when it was not; $F(1,11)=37.580$, $p = .0001$) as well as competitor set (hit rate was .721 in the cohort set, vs. .783 in the rhyme set; $F(1,11)=7.229$, $p = .021$). The interaction was also significant ($F(1,11)=4.966$, $p = .048$). Comparisons of individual means show that within each competitor set, hit rate was significantly lower ($p < .01$) when competition was possible than when it was not (.587 vs. .855 for the cohort competitor set, and .711 vs. .855 for the rhyme competitor set), and that when competition was possible, hit rates were significantly lower when the competitor was a cohort (.587) than when it was a rhyme (.711).

Experiment 2 clearly shows that both cohorts and rhymes compete for lexical activation. Participants were much more likely to launch a "false alarm" eye-movement to a cohort or rhyme than to a non-competitor, as was reflected in significantly lower hit rates when competitors were present. The results also reveal clear differences between cohort and rhyme competition: cohort activation (as measured by probability of false alarm eye-movements) rises more rapidly and has a higher peak than rhyme activation (as predicted by our TRACE simulations), and the presence of a cohort competitor reduces hit rate significantly more than does the presence of a rhyme competitor.

General Discussion

The research presented here demonstrates that eye-movements can provide a remarkably sensitive measure of the time-course of lexical activation during word recognition in continuous speech. We replicated the finding from gating and cross-modal priming that a cohort of words sharing the same onset is temporarily activated as a word unfolds. In addition we found clear evidence for activation of phonologically similar competitors that do not share onsets. Although this result is clearly predicted by continuous activation models of word recognition, it has proved difficult to find supporting evidence with other paradigms.

We also addressed two crucial methodological issues with the eye-tracking paradigm. First, we showed that the use of a restricted set of lexical possibilities does not appear to artificially inflate similarity effects. In particular, no evidence for rhyme effects was found with gating, a task that emphasizes word initial information. Second, we provided clear evidence in support of a simple linking hypothesis between activation levels and the probability of fixating a target. We assumed that the probability of directing attention to a target, and thus making an eye-movement to it, was a direct function of its probability as quantified by applying the Luce choice rule to activation levels computed using the TRACE model. The predicted probabilities closely corresponded to the behavioral data.

The close match between predicted and actual fixation patterns provides strong support for activation models which continuously map input onto a set of lexical candidates. The sensitivity of the response measure and the presence of a clear liking hypothesis between lexical activation and eye-movements suggests that this methodology will be invaluable in exploring even the finest grained questions about the mechanisms underlying spoken word recognition.

Acknowledgments

Supported by NIH resource grant 1-P41-RR09283; NIH HD27206 to MKT; NIH F32DC00210 to PDA, and an NSF Graduate Research Fellowship to JSM.

References

Goldinger, S.D., Luce, P.A., & Pisoni, D.B. (1986). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28, 501-518.

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71-102.

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions during word-recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576-585.

Marslen-Wilson, W., Moss, H.D., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1376-1392.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

McKoon, G., Allbritton, D., & Ratcliff, R. (1996). Sentential context effects on lexical decisions with a cross-modal instead of an all-visual procedure. *Journal of Experimental Psychology: Memory and Cognition*, 22, 1494-1497.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52, 189-234.

Tanenhaus, M.K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J.C. (1995). Integration of visual and linguistic information is spoken-language comprehension. *Science*, 268, 1632-1634.

Zwitserlood, P. (1989). The locus of the effects of the sentential-semantic context in spoken word processing. *Cognition*, 32, 25-64.

Zwitserlood, P. (in press). Cross-modal Priming. *Language and Cognitive Processes*.