JOURNAL OF MEMORY AND LANGUAGE **38,** 419–439 (1998)
ARTICLE NO. ML972558

Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models

Paul D. Allopenna, James S. Magnuson, and Michael K. Tanenhaus

University of Rochester

Eye movements to pictures of four objects on a screen were monitored as participants followed a spoken instruction to move one of the objects, e.g., ''Pick up the beaker; now put it below the diamond'' (Experiment 1) or heard progressively larger gates and tried to identify the referent (Experiment 2). The distractor objects included a cohort competitor with a name that began with the same onset and vowel as the name of the target object (e.g., *beetle*), a rhyme competitor (e.g. *speaker*), and an unrelated competitor (e.g., *carriage*). In Experiment 1, there was clear evidence for both cohort and rhyme activation as predicted by continuous mapping models such as TRACE (McClelland and Elman, 1986) and Shortlist (Norris, 1994). Additionally, the time course and probabilities of eye movements closely corresponded to response probabilities derived from TRACE simulations using the Luce choice rule (Luce, 1959). 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, but no suggestion of rhyme effects. Given that the same sets of pictures were present during the gating task as in Experiment 1, we conclude that the rhyme effects in Experiment 1 were not an artifact of using a small set of visible alternatives.

Current models of spoken word recognition assume that listeners evaluate the unfolding speech input against an activated set of lexical candidates which compete for recognition. Compelling evidence for these assumptions comes from studies demonstrating that the recognition time for a spoken word is strongly influenced by the set of words to which it is phonetically similar (for a recent review see Cutler, 1995). For example, the recognition time for polysyllabic content words is correlated with the point in the speech stream where the phonetic information becomes consistent with only a single lexical candidate (Tyler, 1984). In addition, recognition time for spoken words is affected by the number and frequency of other words that differ by only a single phoneme (Goldinger, Luce, & Pisoni, 1989; Luce, Pisoni, & Goldinger, 1990).

Although it is clear that multiple candidates compete for recognition, it is less clear just how the competitor set is defined and how it is evaluated. Models of lexical access make different claims about the nature of the competitor set and about how tolerant the processing system is to phonological mismatches between the incoming speech and potential lexical representations.

In the cohort model developed by Marslen-Wilson and colleagues (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978), the onset of a word activates a set of lexical candidates which together comprise a cohort which competes for recognition. For example, as the word *beaker* is presented to the processing system, both *beaker* and *beetle* would initially become active members of the recognition cohort. The activation of members is reduced

We thank Delphine Dahan, Kathy Eberhard, Julie Sedivy, and Michael Spivey-Knowlton for helpful suggestions, and Cynthia Connine, Steve Goldinger, and Dennis Norris for comments which substantially improved this article.

We also thank the UCSD Center for Research in Language for making TRACE freely available and Howard Nusbaum for the use of the Wordprobe utility.

Supported by NIH HD27206 to MKT, NIH F32DC00210 to PDA, and an NSF Graduate Research Fellowship to JSM.

Address all correspondence to Paul D. Allopenna, Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, New York 14627. E-mail may be sent to allopen@bcs.rochester.edu.

0749-596X/98 $25.00

when mismatches are detected over time between candidate representations and the continuing speech input. Thus, the activation of *beetle* would begin to decline at the second consonant of *beaker,* because the input is no longer consistent with its lexical representation. Selection occurs when the evidence is sufficiently strong to support one of the alternatives.

Extensive empirical evidence supports the claim that words sharing initial segments are briefly activated together 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 (Marslen-Wilson, 1989; Zwitserlood, 1989). Thus, *beaker* would not only prime *glass,* an associate of *beaker,* but also briefly prime *bug,* an associate of *beetle,* indicating that lexical representations for both words were initially activated. A similar conclusion comes from studies using the gating task (Grosjean, 1980) in which listeners are presented with successively longer fragments of words. Recognition of polysyllabic content words, as determined by accuracy and confidence judgments, occurred shortly after the speech input became consistent with only one lexical alternative, indicating that until that point multiple lexical alternatives had been active (e.g., Tyler, 1984).

However, as it has been pointed out frequently in the literature, the cohort model makes some problematic assumptions (McClelland & Elman, 1986; Norris, 1990). For example, word onsets in continuous speech are often not clearly marked. Thus, it may not be valid to assume that listeners can reliably identify which information begins a word. Moreover, lexical candidates that have only a partial match to the onset of the word will never enter into the recognition set. In noisy environments, a typical situation for speech, this will limit the robustness of the model and require special recovery mechanisms.

Continuous mapping models, such as the TRACE model (McClelland & Elman, 1986) and the Shortlist model (Norris, 1994), address these problems by assuming that lexical access takes place continuously. In these models, the initial portion of a spoken word still exerts a strong influence on which alternatives are activated shortly after the word begins. However, the set of activated alternatives may also include words that do not have the same onset. For example, as a polysyllabic word unfolds, words that rhyme with the input word will gradually become weakly activated. Thus, *beaker* is predicted to activate a rhyme, such as *speaker,* as well as words sharing initial segments, such as *beetle.*

Relaxing the strict sequential constraints imposed by alignment models such as the cohort model has the desirable property of allowing lexical access to be successful without assuming that onsets are clearly marked in the speech or that there is an initial segmentation stage in processing which aligns the recognition mechanism with word onsets. 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. Thus, it allows lexical candidates that do not begin with the same segments to become activated, which may be important for segmentation (McClelland & Elman, 1986; Norris, 1994) under noisy conditions, for example.

However, the empirical evidence for activation of lexical competitors that do not have similar onsets is inconclusive (cf. Zwitserlood, 1996). Several studies have demonstrated priming for embedded words, e.g., *bone* in *trombone* (Shillcock, 1990), as well as potential words when word boundaries are ambiguous (Tabossi, Burani, & Scott, 1995). In addition, work by Luce and colleagues (Luce et al., 1990) demonstrates that recognition time for a word is influenced by the density of its lexical neighborhood, where a neighborhood is composed of similar lexical items and similarity is defined across the entire neighborhood, including words with dissimilar onsets.

The most direct attempts to find evidence for activation of potential lexical competitors differing in their onsets have examined rhyme priming. Although nonword primes do seem to activate rhyming words (e.g., Connine, Blasko, & Titone, 1993), studies using real words as primes find evidence for activation only when the competitor and the target have very similar onsets. Marslen-Wilson and Zwitserlood (1989) failed to find evidence for rhyme priming using a cross-modal lexical decision task, although they reported post hoc analyses which suggested there might have been weak rhyme activation for some of their items. In contrast, Connine et al. (1993) did find evidence for rhyme priming using nonword primes, but only when the input diverged from a base word by one or two features. Andruski, Blumstein, and Burton (1994) reported rhyme effects using prime–target pairs in which the prime word onset was always a voiceless-stop that was sometimes distorted (by reducing the VOT). Lexical decision times to pairs where the prime had a voiced-stop counterpart (e.g., *pat*/*bat*) were slower than they were to pairs for which the prime had no voiced-stop lexical competitor (*king*/*ging*).

Most recently, Marslen-Wilson et al. (1996) found similar results to Connine et al. (1993) using cross-modal priming. However, Marslen-Wilson et al. did not find rhyme priming using an auditory–auditory priming task. Based on a time-course difference between the two tasks they argued that while candidates which are ambiguous at onset can be recognized as a token of a word, this happens only as a result of a late perceptual stage of processing that follows initial contact with the lexicon during ‘‘preperceptual’’ processing. Marslen-Wilson et al. concluded that initial contact with the lexicon excludes candidates that do not share onsets, as predicted by the cohort model, but eventual lexical selection is more error tolerant.

In sum, whereas continuous mapping models clearly predict that potential lexical candidates with mismatching onsets will become activated, empirical studies investigating this prediction using rhymes find that a stimulus will only activate lexical representations of rhymes with closely matching onsets.

The lexical recognition system may, in fact, be as finely tuned to feature mismatches as these studies suggest. However, an alternative possibility is that a lexical competitor that partially matches the input might be weakly activated but not sufficiently so to be detected in a task such as cross-modal priming. That is, semantic priming may be too insensitive or too indirect a response measure to reveal activation of rhymes. This would lead to the erroneous conclusion that the system is more finely tuned to feature mismatches than it actually is, and it would underestimate the effects of lexical candidates that might not become active until relatively late in a word.

In recent research we have been exploring a paradigm in which participants follow spoken instructions to manipulate either real objects or pictures displayed on a computer screen while their eye movements are monitored using a lightweight camera mounted on a headband (Tanenhaus & Spivey-Knowlton, 1996; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995, 1996). We find that eye movements to objects in the workspace are closely time-locked to referring expressions in the unfolding speech stream, providing a sensitive and nondisruptive measure of spoken language comprehension during continuous speech.

The eye movement paradigm could be quite a valuable methodology for studying lexical access in spoken word recognition. Lexical processing can be studied in the context of ongoing comprehension using continuous speech input in fairly natural tasks. In addition, unlike most other methods used for studying spoken word recognition, the paradigm does not involve either interrupting speech or asking participants to make a metalinguistic decision about the input.

Preliminary work suggests that eye movements might provide a level of sensitivity necessary for exploring the time course of subtle competitor effects in spoken word recognition. Spivey-Knowlton and colleagues (Spivey-Knowlton, 1996; Tanenhaus et al., 1995) had participants pick up and move real objects in response to instructions such as: ‘‘Look at the cross [i.e., fixation point]. Now pick up the candle and hold it above the cross. Now put it below the candy.’’ The set of objects sometimes included an object with a name beginning with the same phonemic sequence as the target object (an onset cohort competitor, e.g.

candy and *candle,* or *cart* and *carton;* for details see Spivey-Knowlton, 1996 and Spivey-Knowlton & Tanenhaus, submitted).

The presence of a cohort competitor increased the latency of eye movements to the target and induced frequent looks to the competitor, including trials in which the initial fixation was to the competitor. These results indicated that the two objects with similar names were, in fact, competing as the target word unfolded. The timing of the eye movements also indicated that they were programmed during the ambiguous portion of the target word. A time course analysis of the proportion of fixations on the target object (correct referent), cohort competitor, and unrelated objects suggested that this measure would be extremely sensitive to the uptake of information during lexical access. Furthermore, the shapes of the functions suggested that they could be closely mapped onto activation levels.

The research presented here extended the eye movement paradigm to examine whether competitor effects would be seen for objects with names that rhyme with the target, as predicted by continuous mapping models. We also put forth an explicit account of the mapping between activation levels simulated by the TRACE model and fixation probabilities. Additionally, we evaluated the concern that using a limited set of visual alternatives would artificially inflate similarity effects.

EXPERIMENT 1

This experiment examined the time course of cohort and rhyme activation in order to test predictions made by models such as TRACE and Shortlist, in which the speech input is continuously mapped onto lexical representations. Participants were presented with line drawings of four objects on a computer screen and instructed to move the objects (by clicking on them and dragging them with the computer's mouse) to locations defined with respect to fixed geometric shapes (e.g., ''Pick up the beaker. Now put it above the diamond.''). The displays were generated from eight sets of four objects. Each set had a referent, an onset cohort competitor, a rhyme competitor, and an unrelated distractor. The eight sets are presented in Table 1, along with frequency, familiarity, and neighborhood information.[1]

Simulations with TRACE

In order to confirm that continuous mapping models actually predict activation of rhymes, and to quantify the predicted effects, simulations with the word sets used in the experiment were conducted with the TRACE model developed by McClelland and Elman (1986). TRACE was chosen as an example of a continuous mapping model because it has been well studied and because an explicit computational implementation is available.[2]

TRACE is a hierarchically structured interactive activation network with feature, phoneme and word level units. The network has fixed connections both between and within levels. Connections between levels are bi-directional and excitatory, whereas connections within a level are inhibitory. Processing proceeds by applying an idealized spectral representation to the feature level. Subsequently, there is typically both bottom-up and top-down information flow throughout the system. Between-level excitation serves to activate words that match the input to some degree, and the within-level inhibitory connections implement a competitive mechanism that helps boost the activation of the word that most closely matches the input.

TRACE provides two free input parameters which we set for the simulations reported here. The first is a duration parameter, which determines the relative duration (over discrete time slices) for which each phoneme remains active. We selected the default setting of 1.0. The second parameter determines the strength of the input of each phoneme and can vary (typically between 0 and 1.0). In our simulations, we assumed an input strength of 1.0.

[1] These statistics were retrieved using the Wordprobe utility developed in Howard Nusbaum's laboratory at the University of Chicago. Familiarity is based on the seven-point ratings obtained by Nusbaum, Pisoni, and Davis (1984).

[2] We used the implementation of TRACE available from the UCSD Center for Research in Language via anonymous ftp at ftp://crl.ucsd.edu/pub/neuralnets/.

TABLE 1

Items Used in the First Experiment

Pair	Referent	Cohort	Rhyme	Unrelated
A	beaker (2) (6.7) (4)	beetle (0) (7.0) (6)	speaker (49) (−1.0) (3)	dolphin (1) (7.0) (2)
	carrot (1) (6.9) (7)	carriage (11) (7.0) (3)	parrot (1) (7.0) (9)	nickel (7) (7.0) (8)
B	candle (18) (7.0) (8)	candy (16) (7.0) (5)	handle (53) (7.0) (5)	dollar (46) (7.0) (8)
	pickle (1) (7.0) (8)	picture (162) (6.8) (3)	nickel (7) (7.0) (8)	speaker (49) (−1.0) (3)
C	casket (0) (7.0) (3)	castle (8) (6.6) (11)	basket (17) (7.0) (4)	nickel (7) (7.0) (8)
	paddle (1) (7.0) (9)	padlock (2) (7.0) (1)	saddle (25) (6.7) (5)	dollar (46) (7.0) (8)
D	dollar (46) (7.0) (8)	dolphin (1) (7.0) (2)	collar (17) (7.0) (15)	beaker (2) (6.7) (4)
	sandal (0) (6.6) (7)	sandwich (10) (7.0) (1)	candle (18) (7.0) (8)	parrot (1) (7.0) (9)

Note. Different pairs of sets were presented to different groups of participants. The three numbers given below each word are its frequency (per million words in the Kucera and Francis, 1967, corpus), its familiarity (based on 7-point ratings obtained by Nusbaum *et al.*, 1984; values of −1.0 indicate that the item was not included in the rating study), and a count of its (noun) phonological neighbors.

Each input word was run for 90 cycles. New phonemes were introduced every sixth cycle and the input for each successive phoneme was active for 11 cycles (see McClelland & Elman, 1986, for details).

Each simulation was conducted with a 268 word lexicon. The lexicon included the 230 words provided in the TRACE simulation package along with any experimental items that were not included. In addition, neighbors for all of the words used in the experiment were added to the lexicon if they were not already included. Neighbors were defined as any word that differed from a base word by no more than one phoneme (and were found automatically using the Wordprobe utility). Neighbors were included in order to ensure that the lexicon included representative neighborhoods for the critical items.

Simulations were run using each of the eight referent items in Table 1. The input was the word ‘‘the’’ followed by the referent word. ‘‘The’’ was included as input to simulate the fact that words were presented in continuous speech using instructions with the carrier phrase ‘‘pick up the ____’’. Word boundaries were not marked.

Figure 1 shows the average activation levels from simulations with the referents from the eight stimulus sets. The activation functions were converted into predicted fixation probabilities across time in order to compare predic-

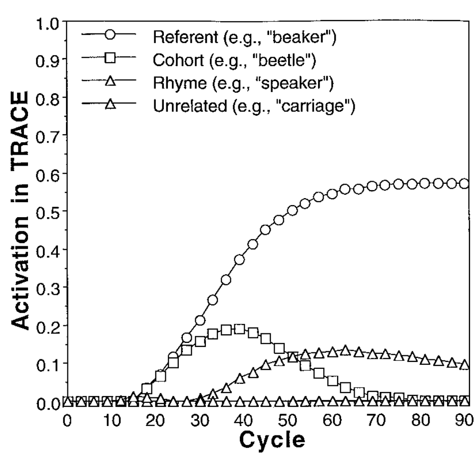

FIG. 1. Average activations from eight TRACE simulations with both cohort and rhyme competitors.

tions from the model with the human data. This required developing an explicit linking hypothesis regarding how activation levels map onto fixations in the task we used. We made the general assumption that the probability of initiating an eye movement to fixate on a target object o at time t is a direct function of the probability that o is the target given the speech input and where the probability of fixating o is determined by the activation level of its lexical entry relative to the activations of the other potential targets (i.e., the other visible objects). This simple linking hypothesis assumes that the probability of directing attention to an object in space along with a concomitant eye movement—either to obtain more information about the object or to guide a hand-movement—is a direct function of the probability that it is the object to be picked up. Note that this hypothesis does *not* require stronger and less defensible assumptions about the relationship between eye movements and attention. For example, we are not committed to the assumption that scan patterns in and of themselves reveal underlying cognitive processes. Nor do we assume that the fixation location at time t necessarily reveals where attention is directed (see Vivianni, 1990 for an extended discussion of these issues). Rather, given carefully constrained tasks and stimuli, fixations are probabilistically related to attention.

Activations were first transformed into response strengths using Eq. [1], following Luce (1959),

$$S_i = e^{ka_i}, \qquad [1]$$

where S and a are the response strengths and activations for each item, i, and k is a free parameter that determines the amount of separation between units of different activations. Explorations with the model indicated that the best fits occurred when k was a sigmoid function (comparisons of fits with sigmoidal and fixed k values are presented below). As time passes, the separation parameter increases. This function is conceptually similar to the dynamic criterion that Spivey-Knowlton and colleagues have used to simulate eye-movement data using an integration competition model (e.g., McRae, Spivey-Knowlton & Tanenhaus, 1998). Fixations are generated when one of several competing interpretation nodes reaches a criterion. As time—measured in processing cycles—passes, the criterion is lowered. In addition to simulations using a sigmoid function, we also report simulations in which k was set to a fixed value.

Response strengths were then converted into response probabilities using the Luce choice using Eq. [2], where L_i is the response probability for item i of j items based on the basic Luce choice rule:

$$L_i = \frac{S_i}{\Sigma S_j}\,. \qquad [2]$$

We made the simplifying assumption that only the activations for those words pictured in the visual display would be evaluated using the choice rule. The rationale for this was that only pictured items were available as possible responses. More generally, though, this raises the question of how the information in the visual display combines with, and interacts with, the speech input. We return to this issue in more detail in the general discussion.

In order to guarantee that the probability of fixating an object was determined both by its activation and by its activation relative to the other alternatives, we transformed the response probability so that it could vary from 0 to 1.0. (The Luce choice rule assumes that prior to any stimulus presentation, all responses are equally likely; i.e., for j items, the minimum response probability is $1/j$.) Thus, a scaling factor was calculated at each time slice, t:

$$\Delta_t = \frac{\max(act(t))}{\max(act(overall))}\,. \qquad [3]$$

Equation [3] yields a scaling factor based both on the activation at the current time slice ($\max(act(t))$, the maximum activation of the four objects at time t) and the activation levels observed throughout the simulation (max (*act*(*overall*)), the maximum activation observed at any time slice). A scaled response probability was generated by multiplying the response probability from the Luce choice rule by the scaling factor

$$p(R_i) = \Delta_t L_i. \qquad [4]$$

Initially as the word unfolds, activations will be small, resulting in a low probability of generating an eye movement to fixate on any of the objects. As more of the word is processed activation increases, resulting in a greater probability of making an eye movement.

Cycles in TRACE were equated with real time by first measuring the mean duration for the words used to refer to the targets (375 ms). This duration was divided by the mean number of phonemes in the target words to yield a measure of the mean number of milliseconds per phoneme. Then, the mean number of cycles per target word, as represented in TRACE, was computed and divided by the mean number of phonemes. This leads us to equate 1 cycle in TRACE to approximately 11 ms of real time. Because the sampling rate of the video record of the experiment was 30 Hz, this meant three cycles in TRACE corresponded to one 33 ms video frame. Therefore, activations from TRACE were recorded every third cycle.

Finally, the model as described thus far does not assume any delay between when a fixation is predicted to occur and the response probability associated with the activation of a candidate at a particular point in the speech. However, it is well established that there is at least a 150-ms delay between when a saccadic eye movement is programmed and when a fixation occurs in much simpler tasks than the one used here (e.g., Matin, Shao & Boff, 1993). Therefore, we added six cycles with zero activation. This value was chosen so that the first predicted fixations occurred at 200 ms. This delay generated the best fits between predicted fixations to the referent and the actual data. By fixing the delay for only the referent, we were able to see how well the model fit both the rhyme and cohort data without explicitly aligning their respective rise times. Figure 2 shows the predicted fixation probabilities averaged across the eight stimulus sets listed in Table 1 using a sigmoid function for k described in Eq. [5], where x is a step increment that determines the steepness of the function, and C is a constant.

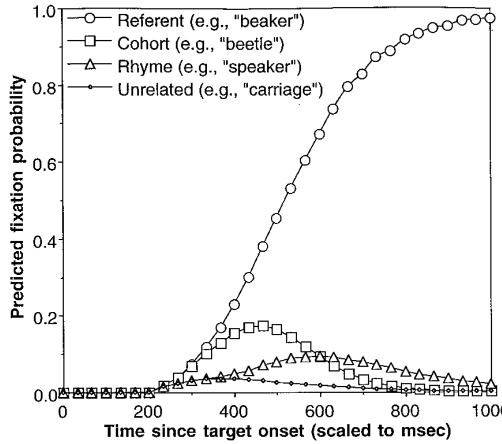

FIG. 2. Predicted response probabilities converted from TRACE using the scaled Luce choice rule.

$$k = \frac{1}{1 + e^{-x}} C. \qquad [5]$$

As the input unfolds, the predicted probability of fixating the referent and the cohort competitor increases relative to those for the rhyme and unrelated items. Slightly later in the speech stream, but before the cohort and referent are predicted to diverge, the probability of fixating the rhyme competitor increases. The probability of fixating the rhyme is never as high as the probability of fixating the cohort competitor, although it is greater than the probability of fixating the unrelated item. Moreover, the probability of fixating the rhyme late in the word is greater than the probability of fixating the cohort. These predictions arise because as the speech unfolds over time, the input becomes increasingly similar to the representation of the rhyme and dissimilar to the representation of the cohort. The rhyme activation remains relatively low because by the point where the input becomes consistent with the rhyme, the referent is already highly activated.

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 – unrelated'' sets presented in Table 1. The sets were divided into four pairs (labeled A–D in Table 1), which were presented to different groups of participants. On any trial, four line drawings were presented to participants on a computer display, and the participants were instructed to click on and move one of the objects using a computer mouse. There were four possible combinations of objects: a *full competitor* set, consisting of a referent, a cohort and rhyme, and one unrelated object (e.g., *beaker, beetle, speaker,* and *carriage*); a *cohort competitor* set, consisting of a referent, a cohort, and two unrelated objects (e.g., *beaker, beetle, parrot,* and *carriage*); a *rhyme competitor* set, consisting of a referent, a rhyme, and two unrelated objects (e.g., *beaker, speaker, dolphin, carriage*); and an *unrelated* set, consisting of a 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).

Within each competitor set, each item was used as the target an equal number of times. In all sets besides the full competitor set, more than one unrelated item was used, and therefore there were more trials with unrelated targets in those conditions. The conditions, their frequencies, and the items comprising them are shown in Table 2. In addition to making sure all items appeared equally as often as targets (in order to preclude frequency-based strategies by participants), the overall frequency of each item was also controlled. Each item appeared 48 times (the frequencies sum to 96 in Table 2 because one pair of sets shown in Table 1 was presented to each subject, such that, e.g., each referent appeared 48 times for a total of 96). This was achieved by using items from one set of items from a pair as unrelated items for the other set. For example, if pair A from Table 1 was being used, *carriage, parrot,* and *nickel* were used as additional unrelated items for the *beaker–beetle–speaker–dolphin* set.

The stimuli were read from a script by one of the experimenters (PDA). We did not use digitized speech because the available software did not permit us to synchronize the presentation software, the auditory presentations, and the VCR (a solution is under development). However, other experiments using the visual world paradigm (but not a computer display) that were originally conducted using instructions read from a script have since been replicated using digitized speech (e.g., Spivey-Knowlton, 1996). In order to prevent experimenter bias, PDA could not see the subject's display and only had access to the instructions to be read. Crucially, while he knew what the target was because it was mentioned in the ''pick up'' instruction, he did *not* know which other objects were displayed on that trial and so did not know if a given trial was a critical trial or what the competition condition might be.

Procedure

Participants were seated at a comfortable distance from the experimental control computer. Prior to the experiment, participants were twice shown pictures of the stimuli they were to see in the experiment. First they were shown a grid which contained all of the items

TABLE 2

Conditions Used in Experiment 1

Competitor set	Condition	Trials	Target	Distractors		
Full	12	6	referent	cohort	rhyme	unrelated
	11	6	cohort	referent	rhyme	unrelated
	10	6	rhyme	referent	cohort	unrelated
	9	6	unrelated	referent	cohort	rhyme
Cohort	8	6	referent	cohort	unrelated	unrelated
	7	6	cohort	referent	unrelated	unrelated
	6	12	unrelated	referent	cohort	unrelated
Rhyme	5	6	referent	rhyme	unrelated	unrelated
	4	6	rhyme	referent	unrelated	unrelated
	3	12	unrelated	referent	rhyme	unrelated
Noncompetitor	2	6	referent	unrelated	unrelated	unrelated
	1	18	unrelated	referent	unrelated	unrelated

that they were to see during the experiment. These items were each named by the experimenter. Subsequently, they were again shown the grid. During the second viewing, participants were asked to name each of the objects aloud. If the participant incorrectly named an object, they were corrected by the experimenter and shown the object again. With one exception, participants correctly named all of the stimuli on their first attempt.

Eye movements were monitored using an Applied Scientific Laboratories E4000 eye tracker. Two cameras mounted on a lightweight helmet provided 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 position 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. Auditory stimuli were read aloud (as described above). A microphone connected to the HI-8 VCR provided an audio record of each trial.

The structure of each trial was as follows. First, a 5 × 5 grid with nine crosses on it (for calibration purposes) appeared on the monitor. Then, line drawings of the stimuli appeared on the grid, with a cross in the center cell. A schematic of the grid with the pictures from a full competitor set with *beaker* as the referent is presented in Fig. 3. The line drawings for each trial were placed in the cells on the grid that were directly adjacent to the center cross so that each would be an equal distance from the fixation cross. Each cell in the grid was approximately 5 × 5 cm. Participants were seated about 57 cm from the screen. Thus, each cell in the grid subtended about 5 degrees of

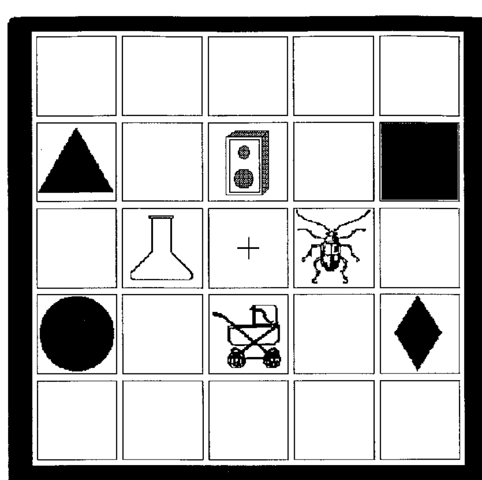

FIG. 3. An example of a stimulus display presented to participants.

visual angle, which is well within the resolution of the tracker (better than 1 degree).

Approximately 2 s after the line drawings appeared, the experimenter instructed the participant to look at the center cross. Prior to the first trial, participants were told they could move their eyes freely until this instruction, but then were to fixate the cross until the next instruction. After approximately 1 s, the experimenter instructed the participant to pick up one of the objects (e.g., ‘‘pick up the beaker’’). Once the participant had clicked on the object with the computer mouse (to pick it up), 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’’). When the subject clicked on the object, it was ‘‘picked up,’’ and moved when the subject moved the mouse. The subject could therefore drag the object to an appropriate location on the screen and then click again to ‘‘drop’’ the object. Once the participant had placed the object in the appropriate square, the experimenter again instructed the participant to look at the center cross. When the participant was looking at the cross—as signaled by clicking on it with the mouse (and verified by a second experimenter monitoring the participant’s fixations)—the next trial began. The grid was then replaced by a blank white screen followed by the calibration screen. Between trials, participants could take a break if they wished. Calibration was monitored by the second experimenter and adjusted between trials when necessary.

Results

The data were analyzed from the videotape records using an editing VCR with frame-by-frame controls and synchronized video and audio channels. Fixations were scored by noting which grid the participant was fixating, beginning with the first fixation after the onset of the target word and ending with the fixation prior to the participant moving the mouse to the correct item. Trials were not included in the analyses if (a) the calibration became so degraded during a trial that fixations could not be reliably coded (a rare event, typically caused by a participant inadvertently bumping the eye tracker), (b) the participant did not maintain fixation on the cross until the appropriate instruction began, or (c) the participant never fixated the correct target. Of 1152 trials, 42 (3.6%) were not included in the analyses. These trials were evenly distributed across conditions. The mean duration of the target words from onset to offset was 375 ms. Scoring began with the frame on which the target word in the instruction began and continued until the fixation prior to pick-up with the mouse.

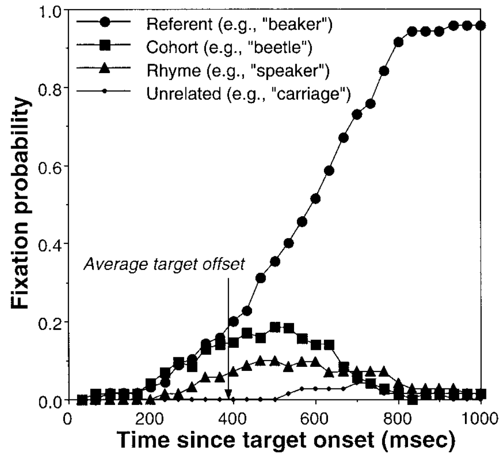

FIG. 4. Probability of fixating each item type over time in the full competitor condition in Experiment 1. The data are averaged over all stimulus sets given in Table 1; the words given in the figure are examples of one set.

Figure 4 presents the fixation probabilities over time in 33-ms intervals (the sampling rate of the video tape record) for the trials on which the referent was presented with a cohort and/or rhyme competitor. Fixations to the referent were averaged across the full competitor trials and the cohort-only and rhyme-only competitor trials. Fixations to the cohort objects were averaged across the full competitor and cohort-only conditions and fixations to the rhyme were averaged over the full competitor and rhyme-only trials. The probabilities do not sum to 1 because the probability of fixating the cross is not plotted.

Participants began fixating on the referent and cohort objects more often than unrelated objects beginning about 200 ms after the onset of the target word, with the referent diverging from the cohort at about 400 ms. Fixations to rhymes began shortly after 300 ms. Fixations to the cohort began earlier and have a higher peak than fixations to the rhyme, but rhyme fixations continued longer relative to the baseline. The general patterns of fixation probabilities clearly follow the same general patterns predicted from the simulations using TRACE. Figure 5 shows the model predictions along with the fixation data for the critical items; the referent and cohort are presented in the upper panel, and the referent and rhyme are presented in the lower panel. Every cycle of TRACE corresponded to 11 ms and we sampled from TRACE every three cycles. Thus, each sample corresponded to one 33-ms video frame.

In order to quantify the goodness of fit between the predicted fixations and behavioral data we calculated the root mean squared (RMS) error for the predicted fixations from the model on the referent, cohort, rhyme, and unrelated objects every three cycles for 30 intervals and the participants' fixation probabilities for the first 30 frames (by which point the probability of fixating the referent had nearly always reached 1.0). Thus, 120 data points were used in the analysis. The overall RMS error was .03, indicating a close fit between the model predictions and the data. A regression comparing the model's predicted fixations with the actual fixations resulted in an r^2 of .99. Table 3 presents RMS and r^2 values for each of the three related item types using a sigmoid function for k and also a fixed value of k (7). As the table shows, the model generated good quantitative fits to the data for all of the critical conditions.

Planned comparisons (one-tailed t-tests) were conducted over response rate (i.e., the probability of fixating each item, inclusive of all fixations) in the various competitor conditions. This was done in order to determine whether the probability of fixating a cohort or a rhyme competitor was reliably greater than the probability of fixating an unrelated object. For those trials on which there was a cohort competitor present (the six full competitor and six cohort competitor trials), participants were more likely to fixate the cohort objects than the unrelated objects, $t(11) = 8.99$, $p = .0001$. The comparisons were also reliable when only trials on which the first fixation was to the cohort competitor were considered, $t(11) = 5.35$, $p = .0001$. Separate analyses for the full competitor sets and the cohort-only competitor sets also revealed significant cohort effects both for all fixations and for first fixations ($p < .05$).

For those trials on which there was a rhyme competitor present (the full competitor and rhyme-only competitor sets) participants were more likely to fixate the rhyme than the unrelated objects, $t(11) = 2.68$, $p = .0106$ for all fixations, and for first fixations only, $t(11) = 2.14$, $p = .0276$. Separate analyses on the full competitor and rhyme-only competitor sets also revealed significant effects for all comparisons except the first-fixation analysis for the rhymes in the full competitor condition, which was only marginally reliable, $t(11) = 1.52$, $p = .078$.

In order to provide a more fine-grained description of how competition emerged over time, we conducted separate analyses on the first eight 100-ms (three frame) windows. These analyses were conducted on trials in which a cohort competitor was present (full competitor and cohort-only combined) and on trials in which a rhyme competitor was present (full competitor and rhyme-only combined). All comparisons were reliable at $p < .05$ unless otherwise indicated below.

For the 0 to 100-ms interval there were no differences among any of the items. During the second 100-ms interval (100–200 ms), there was a trend toward more fixations to the referent and cohort items compared to the unrelated item, $t(11) = 1.69$, $p = .0593$, and $t(11) = 1.43$, $p = .09$, respectively. More fixations to the referent and the cohort items occurred in the 300- to 600-ms intervals compared to the unrelated item, with the referent separating from the cohort in the 400- to 500-ms interval and the cohort becoming indistinguishable from the unrelated condition at 700 ms ($p > 0.1$).

The rhyme first differed from the unrelated item at the 300- to 400-ms interval. During

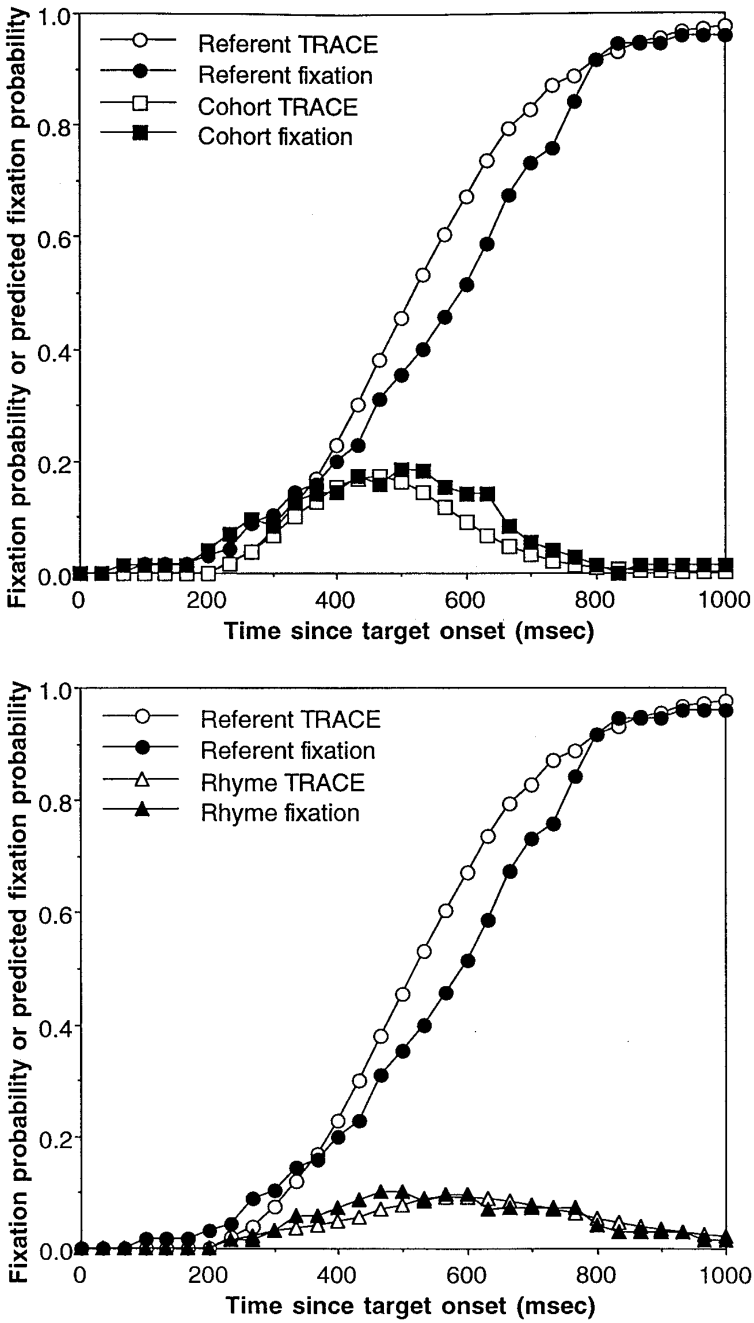

FIG. 5. Model predictions and data for Experiment 1. Predictions and data are shown for the referent and cohort items in the upper panel and for the referent and rhyme in the lower panel.

this interval, fixations to the cohort and referent were more likely than fixations to the rhyme. Beginning at the 400- to 500-ms interval the probability of fixating the cohort no longer differed reliably from the probability of fixating the rhyme, although there were still more fixations to the cohort than the rhyme. Starting with 600- to 700-ms window, there

TABLE 3

Model Fits in Experiment 1 Using Sigmoidal and Fixed k Values for the Related Items

k	Error measure	Referent	Cohort	Rhyme
Sigmoid	RMS	0.07	0.03	0.01
	r^2	0.98	0.90	0.87
Fixed (7)	RMS	0.14	0.05	0.04
	r^2	0.93	0.68	0.57

were more fixations to the rhymes than the cohorts, but this difference never reached significance.

Discussion

The results clearly show that both cohorts and rhymes compete for lexical activation. Participants were more likely to launch an eye movement to a cohort or rhyme than to a noncompetitor. Thus, potential lexical candidates that are sufficiently similar to a spoken word can become activated enough to compete for recognition, whether or not they share the same onset. Whereas rhyme effects have been documented before, they have only been observed in *words* in conditions where there was at most a one-feature difference between the target and a rhyme competitor (e.g., Marslen-Wilson et al., 1996; see Connine et al., 1993, for evidence of rhyme activation given one- or two-feature differences when nonwords were used). In contrast, all of the rhyme competitors used in Experiment 1 differed by more than one feature.

The results also revealed differences between cohort and rhyme competition. Cohort activation, as measured by the proportion of looks to the cohort competitor, rose more rapidly and had a higher peak than rhyme activation. This pattern of eye movements closely matched the patterns predicted by the simulations using TRACE. Overall, the data provide strong support for the hypothesis that speech input is continuously mapped onto potential lexical representations as it unfolds over time.

It is important to note that although the TRACE simulations provided good fits to the behavioral data, the results should be taken as evidence in support of the class of continuous mapping models, rather than support for the particular architectural assumptions made by TRACE. We did not, for example, evaluate the importance of particular features of the TRACE architecture such as lateral inhibition among lexical competitors or lexical/phonemic feedback. We also did not evaluate specific representational assumptions made by TRACE (e.g., the set of features used in TRACE or the assumption of a phonemic level). These are all important questions that remain for future research. Fortunately, the sensitivity of the eye movement patterns to competitor effects, and the success of simulations using a simple linking hypothesis to generate quantitative predictions, suggest that questions of this grain will be amenable to empirical tests.

Earlier, we argued that research using cross-modal semantic priming (as well as other techniques) may underestimate the extent to which rhyme competitors are activated. However, it is possible that the eye movement paradigm could overestimate the extent to which rhymes are activated. One concern is that because we used a relatively small set of pictures, participants might have become aware of the similarity among the referent–cohort–rhyme sets despite the large number of filler trials. Thus, the rhyme effects we observed might have reflected strategies adopted by participants as they became familiar with the stimuli. In order to address this concern, we examined the results using only the critical trials from each subject that occurred early in the experiment.

The analyses were restricted to critical trials which occurred in the first 24 trials of the experiment. Because trials were randomly ordered, of the 576 possible trials (24 trials $\times$ 12 participants), only 47 were critical trials. Thus, the results are based an average of 3.92 critical trial presentations to each subject.

In Fig. 6, fixation probabilities to the critical items are shown for the trials in which the referent was presented with a cohort and/or a rhyme competitor. The figure clearly shows that the same patterns observed in the overall analyses were present in the earliest critical trials.

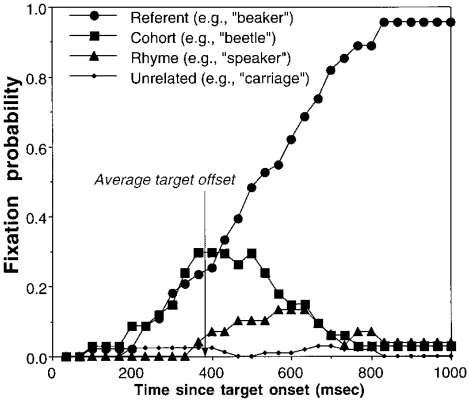

FIG. 6. Probability of fixating each item type over time in the full competitor condition in Experiment 1 in the first quarter of the experiment (the first 24 trials), during which subjects were presented with an average of 3.92 critical trials.

A second concern is that using a restricted set of alternatives might lead to task-induced consideration of any potential target that bears some degree of phonetic similarity to the target word. There is no single experiment that can fully resolve this concern. However, concerns about artifactual effects due to a restricted set would be alleviated if rhyme effects did *not* occur with a restricted set of alternatives under conditions where the presentation conditions emphasized word onsets and thus would not be expected to activate rhymes as strongly as under more normal presentation conditions. Experiment 2 addressed this issue.

EXPERIMENT 2

The first experiment established that words that do not share onsets (or even words that do not align at onset) can nevertheless compete for activation during lexical access. As mentioned earlier, previous research has failed to find consistent competitor effects for lexical items that do not share onsets. There are two possible reasons for this. First, as mentioned above, it is possible that the methodologies used in previous research tend to underestimate rhyme effects, either due to the presentation mode (e.g., either cross-modal priming or gating) or because the test used is not sufficiently sensitive to detect relatively subtle and transient effects. A second possibility is that the rhyme effects seen in the first experiment arose largely as a result of using a restricted set of visible alternatives. The constrained environment used in the experiment may have led to inflated estimates of similarity between all available alternatives.

The goal of this experiment was to determine whether rhyme effects would still be observed with the same display conditions used in Experiment 1, but under auditory presentation conditions which emphasized word-initial information, and thus should lead to weaker rhyme effects. In order to do this we used a gating task with the same stimuli and displays as were used in Experiment 1. Participants heard successively longer fragments of a word on each gate. Their task was to point to which of the four objects (depicted on a CRT with line drawings) was being named. On critical trials, the pictures included the referent (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., *carriage*). Eye movements were monitored during the experiment.

In gating, cohort, but not rhyme, competitors are typically generated as responses throughout the set of gates (e.g., Tyler, 1984). While this is consistent with predictions from the cohort model, it can be argued that this is also consistent with continuous mapping models. The reason is that gating with successively longer segments places clear emphasis on the beginnings of words.

In order to illustrate this point, we conducted a simulation with TRACE in which we modified the input used in Experiment 1 to simulate conditions in gating. On the first gate, a slightly degraded form of the input was presented. The first consonant and vowel phonemes were partially introduced, i.e., their strength and duration parameters were set to less than 1.0. For subsequent gates, these parameters were increased for information presented on previous gates until they reached 1.0. The logic of this manipulation was that repeated presentation of the same information should make it more discriminable. Addition-

TABLE 4

An Example of How Duration and Strength Patterns Were Used To Simulate Gating

Gate	*b*		*i*		*k*		ə		*r*	
1	0.7	0.8	0.5	0.5	—	—	—	—	—	—
2	1.0	1.0	0.7	0.8	0.3	0.2	—	—	—	—
3	1.0	1.0	1.0	1.0	0.5	0.5	—	—	—	—
4	1.0	1.0	1.0	1.0	0.7	0.8	0.3	0.2	—	—
5	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5	—	—
6	1.0	1.0	1.0	1.0	1.0	1.0	0.7	0.8	0.3	0.2
7	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5
8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.7	0.8
9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

ally, subsequent gates added partial information about new successive phonemes. Table 4 illustrates how the strength of the input was adjusted across gates using this procedure for the target word ''beaker.'' Predicted choices and fixations were generated using the same form of the Luce choice rule described earlier. However, we did not scale the response probabilities as we did in the simulations for Experiment 1, because participants were required to make a response at each gate.

Each gating response of the network represents its output after 75 cycles. As in the first simulations, there were eight referent–cohort–rhyme–unrelated sets used for the simulations. In each case, simulations were conducted using nine gates. Figure 7 shows the predicted response probabilities from the simulations for the nine gates averaged across the eight input sets. The simulation predicts that on the first few gates the cohort and referent will have similar response probabilities, with the referent beginning to diverge from the cohort at the third gate. Rhyme and unrelated responses are predicted to occur only rarely. It is important to note that in the simulations rhyme competitors do receive slightly more activation at longer gates than unrelated targets. However, the activation is never high enough to predict differential fixation probabilities for the rhyme and unrelated targets. The simulation demonstrates that the same class of model that predicted rhyme effects using continuous speech predicts that rhyme effects should not occur under successive gating conditions. Thus, the successive gating paradigm offers a test of whether the use of a limited display inflates rhyme effects. If inflated similarity due to visual presence of phonetically similar objects was primarily responsible for the rhyme effects in Experiment 1, then the use of a restricted set of visual alternatives should result in rhyme effects even under stimulus presentation conditions where they would otherwise not be expected to occur. Thus, more rhyme choices or looks to rhyme competitors compared to the unrelated object with gated presentation would suggest that the presence of a limited set was inflating similarity effects. However, the absence of rhyme effects in gating would provide evidence against this hypothesis and would make

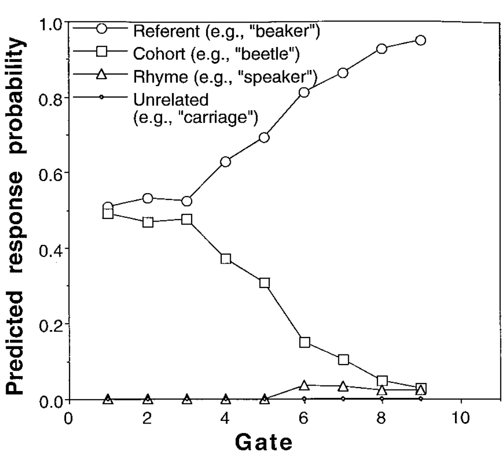

FIG. 7. Predicted probabilities for the gating task using TRACE.

the evidence for rhyme effects with continuous speech found in Experiment 1 more compelling.

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 the eight ''referent–cohort–rhyme'' triples used in Experiment 1. The same grid was used for the stimuli as was used in Experiment 1 except that the geometric shapes were not presented on the display. Stimuli were presented in groups of four, in a 5×5 grid, displayed on a CRT. The center square of the grid contained a cross which the participant was asked to fixate until the presentation of the auditory stimulus. The line drawings for each trial were placed in the cells on the grid that were diagonally adjacent to the center cross.

The auditory stimuli were recorded using the SoundEdit 16 Program. All stimuli were recorded as isolated words which were subsequently spliced into smaller stimuli to create the experimental items. Each token of a stimulus word was recorded at a 44.1 kHz sampling rate with a sampling size of 16 bits. The recorded word tokens were then normalized to have the largest possible gain without clipping.

Altogether, 16 words were presented in the experiment. Eight of the words were presented in critical trials (i.e., trials with both cohort and 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. Each subsequent gate added 40 ms onto the preceding gate.

Procedure

Participants were seated at a comfortable distance from the experimental control 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. If the participant incorrectly named an object, they were corrected by the experimenter and shown the object again. 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 1 s, during which period participants could move their eyes freely. Then, the experimenter instructed the participants to fixate the center cross. 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. Once the auditory stimulus was presented, participants indicated which word they thought they heard by touching the object on the computer screen whose name matched their hypothesis. Eye movements were again monitored using the same procedure as in Experiment 1.

Results and Discussion

Figure 8 shows the probability with which participants selected each of the four pictures for the first eight gates. On the initial gates, participants were equally likely to select the referent and its cohort competitor. On subsequent gates, the probability of se-

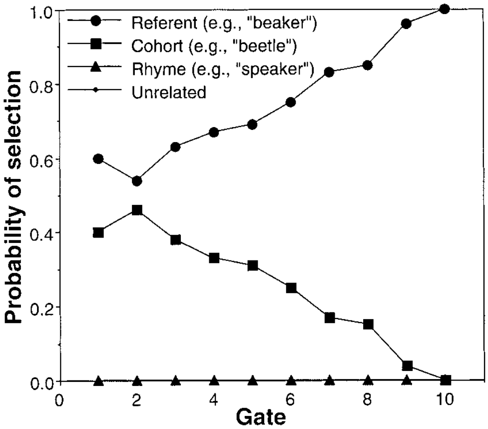

FIG. 8. Probability of selecting each item in Experiment 2.

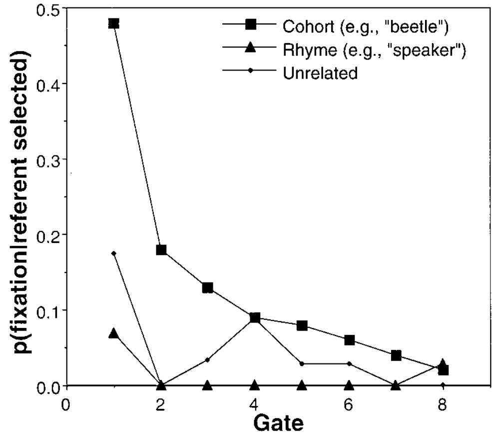

FIG. 10. Conditional probability of an eye movement to either the cohort, rhyme, or unrelated item when the referent was chosen in the gating task.

lecting the referent increased. Rhyme and unrelated objects were never selected and their selection probabilities did not differ from each other. There was a main effect of response type on selection probability, $F(3,15) = 621.66$, $p = .0001$, as well as an interaction between response type and gate, $F(9,45) = 3.05$, $p = .0062$. Comparisons between individual means indicated that the probabilities of selecting the referent and cohort differed reliably beginning at the third gate. The rhymes and unrelated items did not differ reliably at any of the gates.

Figure 9 shows the probability of making an eye movement to each of the pictures across gates. There was a main effect of gate on fixation probability, $F(3,15) = 4.55$, $p = .019$, a main effect of response type ($F(3,15) = 351.27$, $p = 0.0001$), and an interaction between gate and response type, $F(9,45) = 2.97$, $p = .0073$. Comparisons between individual means showed no significant differences between the probability of fixating the referent and the cohort competitor until the fourth gate. Participants rarely looked at either the unrelated or the rhyme items, and the probabilities of fixating these items did not differ. Separate ANOVAs, along with planned comparisons conducted at each gate indicated that fixation probabilities to the items in the rhyme and unrelated conditions did not differ at any of the gates.

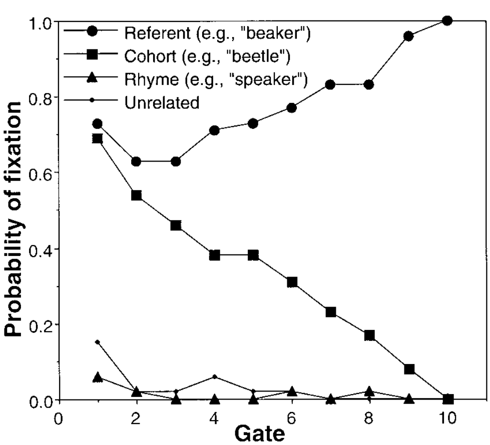

FIG. 9. Probability of fixating each item in Experiment 2.

Figure 10 shows the probability of making an eye movement to the cohort, rhyme and unrelated items when the referent was chosen. The high probability of fixations to the cohort confirms that it was being considered even when the referent was chosen. There was no suggestion that the rhyme was more likely to be fixated than the unrelated object.

The upper panel of Fig. 11 shows that the response probabilities generated from the TRACE simulations provide a good overall fit to the fixation probabilities across successive gates. The filled symbols plot the fixation data. The open symbols show the response probabilities generated from the TRACE sim-

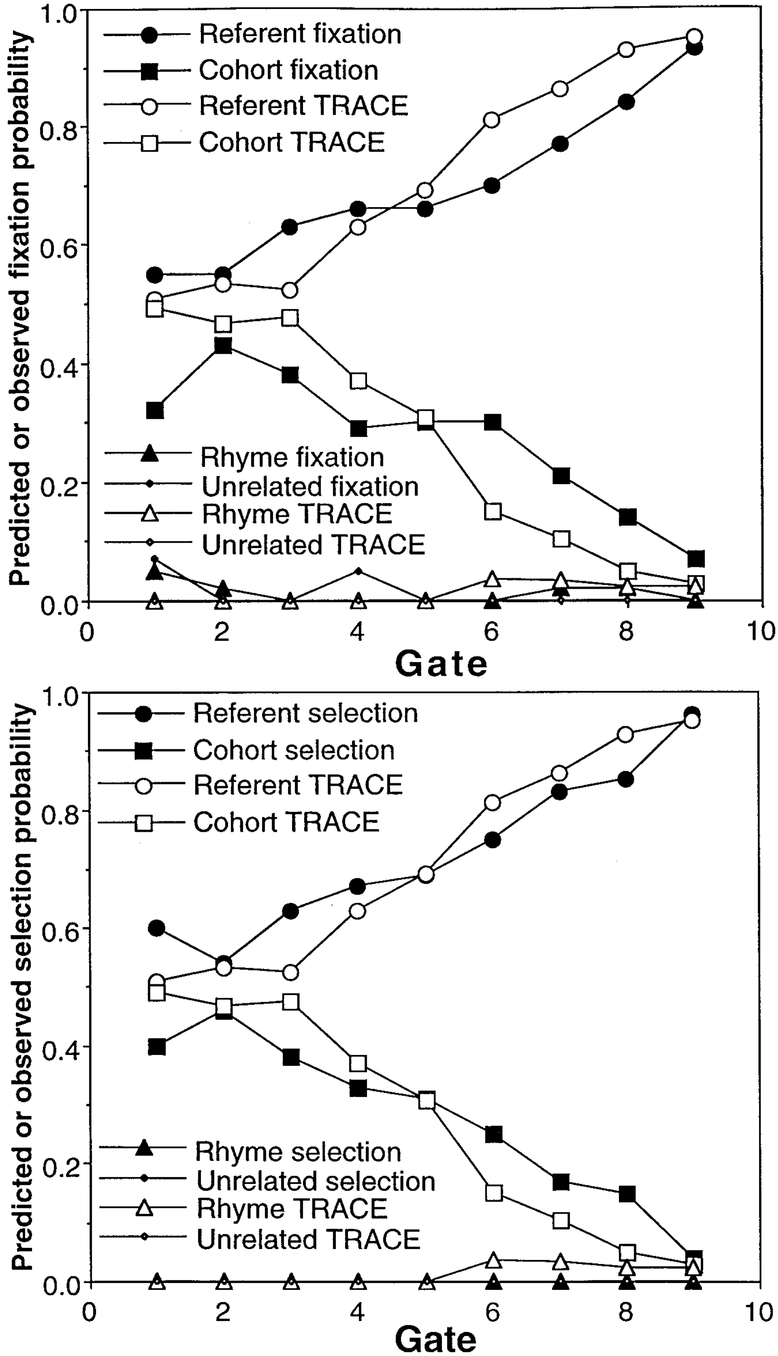

FIG. 11. Model predictions and data for Experiment 2. Predictions are compared with selection data in the lower panel. Predictions are compared with probabilities based on first fixations in the upper panel (see text for details).

ulations. The lower panel of Fig. 11 shows that these response probabilities also predict the probability of selecting each of the four objects across successive gates. The open symbols show the same response probabilities generated from TRACE shown in the up-

TABLE 5

Model Fits in Experiment 2 for the Related Items

Data	Error measure	Referent	Cohort	Rhyme
Selections	RMS	.06	.07	.02
	r^2	.91	.91	.00
Fixations	RMS	.07	.10	.02
	r^2	.88	.77	.02

per panel. Note that the fixation data plotted here is not the same as that presented in Fig. 9, which is clearly less similar to the predicted response probabilities. The fixation data plotted in Fig. 9 included all fixations within a trial. In some cases, participants fixated both the referent and the cohort before making a selection. For the original analysis, we calculated the conditional probability that an eye movement was made to a particular object on a particular gate. By this measure, the conditional probability of looking at any one object is independent of the conditional probability of looking at any other object. In order to compare the experimental results with TRACE—where probabilities must always sum to 1.0—we have included only the first fixation data in Fig. 11. The overall root mean squared error and r^2 values comparing the model predictions to selection data were .05 and .98, respectively, and .05 and .97 for the fixation data. Root mean squared error and r^2 values comparing the model with the observed data for the critical items are presented in Table 5.

Overall, the results provide clear evidence for activation of cohort competitors. In contrast, no evidence was found for activation of potential rhyme competitors even though the set of alternatives was restricted. The fact that there were no rhyme effects with gating—a task that emphasizes word-initial information—reduces concerns that the rhyme effects found in Experiment 1 were an artifact of using a limited set of alternatives. Of course these results do not allow us to definitively conclude that the restricted set did not contribute to the results of Experiment 1. However, any effects would likely be quite small. Under conditions where the model predicted that the similarity of the speech input to the rhyme, relative to the other alternatives, was not sufficient to generate looks to the rhyme target, participants were no more likely to look at the rhyme than at the unrelated item.

GENERAL DISCUSSION

The research presented here provides the first clear evidence for activation of lexical representations of rhyme competitors that differ from an input word by more than one feature at word onset. Although rhyme competition is clearly predicted by continuous activation models of word recognition, it has proved difficult to find supporting evidence with paradigms such as cross-modal priming. These results strongly support predictions made by continuous mapping models. They also suggest that studies using priming methodologies may have overestimated the extent to which the recognition system is sensitive to feature mismatches. Moreover, the time course of the rhyme effects we observed clearly indicates that the effects are a result of the same processes that result in activation to lexical candidates that have similar onsets. Thus, we find no support for the distinction between preperceptual and perceptual stages of lexical processing proposed by Marslen-Wilson et al. (1996).

We also addressed two important 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 successive gating, which is 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 on a target. We assumed that the probability of making an eye movement to a target was a direct function of its predicted response probability which was derived by applying the Luce choice rule to activation levels computed using the TRACE model. The predicted probability that an object would be fixated over time closely corresponded to the behavioral data.

The availability of a mapping between hypothesized activation levels and fixation probabilities that can be used to generate quantitative predictions means that eye movement data can be used to test detailed predictions of explicit models. The sensitivity of the response measure coupled with a clear linking hypothesis between lexical activation and eye movements indicate that this methodology will be invaluable in exploring questions about the microstructure of lexical access during spoken word recognition. The eye movement methodology should be especially well suited to addressing questions about how fine-grained acoustic information affects word recognition. A particularly exciting aspect of the methodology is that it can be naturally extended to issues of segmentation and lexical access in continuous speech under relatively natural conditions.

However, there are important methodological and theoretical concerns about the eye-movement paradigm that will need to be addressed in future work before the full potential of the paradigm can be realized. The most complex issues concern how the presence of a circumscribed visual world interacts with the word recognition process. We made the simplifying assumption that the alternatives act as a filter (in that only the activations of potential responses are included in the choice rule) without affecting the patterns of lexical activations themselves. It will be important in future work to evaluate the degree to which this assumption is viable, and if it needs to be modified, how the modifications could limit the generality of conclusions from the visual world paradigm.

Ultimately, it will be important to have a more comprehensive model that combines a model of activation of lexical alternatives with a model of how activated lexical information is used to identify objects in the world. From informal observation, we strongly suspect that visual similarity among alternatives is an important variable. We also know from pilot work conducted by Spivey-Knowlton that basic cohort competitor effects are found even when the participant has had no prior exposure to the set of alternatives or their names. In this work, objects were placed on a workspace while the participant's eyes were closed. The participant then followed a sequence of instructions such as, ''open your eyes and look at the cross; now pick up the candle.'' In addition, participants in experiments like the ones presented here and participants in experiments with real objects report that they do not generate names for the objects when the display is presented. These observations suggest that participants are using semantic/perceptual representations that become active as soon as information about lexical alternatives begins to be available. This information is used to help identify possible referents from the visual world. We have just begun to explore the set of issues that need to be addressed in order to develop a model of this process.

REFERENCES

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition,* **52,** 163–187.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language,* **32,** 193–210.

Cutler, A. (1995). Spoken word recognition and production. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Vol. 11. Speech, language, and communication* (pp. 97–136). San Diego: Academic Press.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language,* **28,** 508–518.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics,* **28,** 267–283.

Kucera, F., & Francis, W. (1967). *A computational analysis of present day American English.* Providence, RI: Brown University Press.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122–147). Cambridge, MA: MIT.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York: Wiley.

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

Marslen-Wilson, W. (1989). Access and integration: Projecting sound onto meaning. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.

Marslen-Wilson, W., Moss, H. E., & van Halen, S.

(1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* **22,** 1376–1392.

Marslen-Wilson, W., & 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.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception and Psychophysics,* **53,** 372–380.

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

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling thematic fit (and other constraints) within an integration competition framework. *Journal of Memory and Language,* **38,** 283–312.

Norris, D. (1990). A dynamic-net model of human speech recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives.* Cambridge, MA: MIT.

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

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words* (Research on Speech Perception Progress Report Number 10). Bloomington, IN: Speech Research Laboratory, Department of Psychology.

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives.* Cambridge, MA: MIT.

Spivey-Knowlton, M. J., & Tanenhaus, M. K. (submitted). Integration of visual and linguistic information in resolving temporary lexical and structural ambiguities.

Spivey-Knowlton, M. K. (1996). *Integration of visual and linguistic information: Human data and model simulations.* Unpublished Ph. D. thesis, University of Rochester, Rochester.

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language,* **34,** 440–467.

Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. In F. Grosjean & U. Frauenfelder (Eds.). Special issue of *Language and Cognitive Processes: A guide to spoken word recognition paradigms,* **11,** 583–588.

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,** 632–634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye-movements to study spoken language comprehension: Evidence for visually-mediated incremental interpretation. In T. Inui & J. L. McClelland (Eds.), *Attention and performance* (Vol. XVI, pp. 457–458). Cambridge, MA: MIT.

Tyler, L. (1984). The structure of the initial cohort: Evidence from gating. *Perception and Psychophysics,* **36,** 417–427.

Vivianni, 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. Reviews of oculomotor research V4* (pp. 353–383). Amsterdam: Elsevier.

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

Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes,* **11,** 589–596.

(Received July 31, 1997)
(Revision received December 8, 1997)