Cognitive Psychology **42,** 317–367 (2001)
doi:10.1006/cogp.2001.0750, available online at http://www.idealibrary.com on IDEAL®

Time Course of Frequency Effects in Spoken-Word Recognition: Evidence from Eye Movements

Delphine Dahan, James S. Magnuson, and Michael K. Tanenhaus

University of Rochester

In two experiments, eye movements were monitored as participants followed spoken instructions to click on and move pictures with a computer mouse. In Experiment 1, a referent picture (e.g., the picture of a bench) was presented along with three pictures, two of which had names that shared the same initial phonemes as the name of the referent (e.g., *bed* and *bell*). Participants were more likely to fixate the picture with the higher frequency name (*bed*) than the picture with the lower frequency name (*bell*). In Experiment 2, referent pictures were presented with three unrelated distractors. Fixation latencies to referents with high-frequency names were shorter than those to referents with low-frequency names. The proportion of fixations to the referents and distractors were analyzed in 33-ms time slices to provide fine-grained information about the time course of frequency effects. These analyses established that frequency affects the earliest moments of lexical access and rule out a late-acting, decision-bias locus for frequency. Simulations using models in which frequency operates on resting-activation levels, on connection strengths, and as a postactivation decision bias provided further constraints on the locus of frequency effects.

Key Words: lexical frequency; spoken-word recognition; eye tracking.

As the sound pattern of a spoken word unfolds over time, recognition takes place against a backdrop of partially activated alternatives that compete for recognition. The most activated alternatives are those that most closely match the input. For instance, as a listener hears the word *cap*, lexical representations of words with similar sounds, such as *cat*, will be briefly activated.

This work was supported by NSF Grant SBR-9729095 to Michael K. Tanenhaus and Richard N. Aslin, and an NSF GRF to James S. Magnuson. We thank Ellen Hogan for her help in constructing the materials and Ellen Hogan, Sarah Scheidel, and Dana Subik for their help in testing participants. We thank Paul Allopenna, Mark Pitt, and Uli Frauenfelder for technical advice on the TRACE model. Helpful comments from Fernanda Ferreira led to some of the analyses presented in Experiment 1. We thank her and three anonymous reviewers for suggestions that improved this article.

Address correspondence and reprint requests to Delphine Dahan, Max-Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. Fax: (31) 24 352 1213. E-mail: delphine.dahan@mpi.nl.

0010-0285/01 $35.00

The number of competitors, their frequency of occurrence in the language, as well as the frequency of occurrence of the target word itself all affect recognition (e.g., Luce & Pisoni, 1998; Marslen-Wilson, 1987, 1990). The present study focuses on the time course of frequency effects for words and their close competitors.

A variety of response measures has established that recognition of low-frequency words is poorer than recognition of high-frequency words (e.g., Howes & Solomon, 1951). Superior performance on high-frequency words may be observed either in the speed of the response when the stimulus information is sufficient to support accurate performance, as in most reaction-time paradigms (e.g., Marslen-Wilson, 1987) or in the accuracy of the response when the stimulus is degraded (e.g., Luce, 1986; Goldinger, Luce, & Pisoni, 1989).

In most current models of spoken-word recognition, frequency is assumed to affect the activation levels of competing lexical candidates during lexical access. For example, in models with discrete lexical representations such as the ''Cohort'' model (Marslen-Wilson, 1987) or the TRACE model (McClelland & Elman, 1986), high-frequency words would be processed faster than low-frequency words because frequency determines either the baseline activation level of each lexical unit (McClelland & Rumelhart, 1981; Marslen-Wilson, 1990) or the strength of the connections from sublexical to lexical units (MacKay, 1982, 1987). In distributed learning models, the representations of high-frequency words would be activated more rapidly because high-frequency mappings are better learned, resulting in stronger connection weights (Gaskell & Marslen-Wilson, 1997; Plaut, McClelland, Seidenberg, & Patterson, 1996). In contrast, the Neighborhood Activation Model (henceforth, NAM; Luce, 1986; Luce & Pisoni, 1998) places the locus of frequency effects in a decision stage that follows initial lexical activation. In this model, the input stimulus activates a set of acoustic–phonetic patterns that shares some degree of similarity with the input; these acoustic–phonetic patterns in turn activate word-decision units that are tuned to them. High-level information, such as word frequency, is associated with each word-decision unit. Acoustic–phonetic information is assumed to drive the system by activating word-decision units, whereas high-level lexical information is assumed to operate by biasing these decision units. Frequency bias operates by adjusting the activation levels represented within the word-decision units. Thus, word frequency is not directly coded in the resting-activation level, but rather operates as a bias on the activation of the word-decision unit. Nonetheless, frequency comes into play before lexical access is completed.

An alternative hypothesis about the locus of frequency effects in word recognition has been suggested by Balota and Chumbley (1984, 1985). They argued that the facilitation for responding to high-frequency words compared to low-frequency words in a task such as visual lexical decision mostly reflects a *response* bias that operates at a decision stage. Similarly, word-frequency effects in a pronunciation task (i.e., naming a word visually displayed) are attributed to the word-production process, which is presumed to follow lexical access. This conception of frequency as a response bias appears similar to the decision bias incorporated in the NAM. However, it further assumes that frequency effects occur late in processing, as a response bias that affects decisions after lexical access is completed.

The locus of frequency effects has been widely debated in the visual-word recognition and naming literature (e.g., Balota & Chumbley, 1990; Monsell, Doyle, & Haggard, 1989; see Monsell, 1991, for a review). Paap and colleagues (Paap, Newsome, McDonald, & Schvaneveldt, 1982; Paap & Johansen, 1994) argued that frequency effects occur during a verification phase, following the initial activation phase. However, this claim has been challenged by Allen and colleagues (Allen, McNeal, & Kvak, 1992; Allen, Smith, Lien, Weber, & Madden, 1997). Models adopting a distributed representation of word knowledge, however, do not assume distinct stages during word recognition, such as pre- or postlexical access stages (e.g., Seidenberg & McClelland, 1989; Plaut et al., 1996). Such models show sensitivity to word frequency because words that are presented more often during training have a larger impact on the settings of the connection weights (see McCann & Besner, 1987, for a similar proposal). Thus, frequency can affect multiple levels in the system: In a naming task, it affects both the computation of the phonological code and the conversion of the computed phonological code into a sequence of articulatory-motor commands (see McRae, Jared, & Seidenberg, 1990).

Although subject to intense debate in the visual-word recognition literature, the locus of frequency effects has received less attention in the spoken-word recognition literature. Of particular interest are studies by Marslen-Wilson (1987) and Zwitserlood (1989). In these studies, a spoken-word fragment that matched both a high-frequency and a low-frequency lexical candidate was presented, followed shortly by the visual presentation of a target word; participants were instructed to perform a lexical decision on the visual target. The results revealed less facilitation (compared to a control) for target words semantically associated with the low-frequency candidate than for target words associated with the high-frequency candidate. This was interpreted as evidence that the higher frequency candidate becomes more activated than the lower frequency candidate before the auditory input could distinguish between the two. However, Connine, Titone, and Wang (1993) argued that because the input was incomplete, subjects implicitly completed the input with the more frequent alternative on a greater proportion of the trials; this would lead to a larger priming effect for the more frequently generated alternative. On this account, frequency is being used to bias selection of the higher frequency alternative rather than its activation. The same argument can be applied to studies demonstrating word-frequency effects using the gating paradigm (Grosjean, 1980; Tyler, 1984).

Connine et al. (1993) provided evidence that they interpreted as support for models in which word frequency operates as a response bias without affecting activation levels. In this study, participants heard stimuli with initial stop consonants that varied along a voicing continuum; these ambiguous sequences could map onto a high- or low-frequency word (e.g., *?est*, halfway between *best* and *pest*). Participants' consonant identifications and response times were analyzed. Previous research has shown that consonant identifications are influenced by the frequency of the alternatives (e.g., *?est* is identified as *best* more often than *pest*), revealing lexical influence on phoneme identification (Fox, 1984). Crucially, in the Connine et al. study, the ambiguous stimuli were presented along with unambiguous words of high and low frequencies (mixed list), or with only high-frequency words (high-frequency list), or with only low-frequency words (low-frequency list). The manipulation of extrinsic frequency (i.e., the list bias) was predicted to exaggerate the effect of intrinsic word frequency in the high-frequency list and to reverse the effect of intrinsic word frequency in the low-frequency list. Extrinsic frequency is assumed to operate as a response bias, late in the process. If intrinsic word frequency is coded in resting activation levels, Connine et al. reasoned, its influence should still be observed at fast response times, before extrinsic frequency comes into play (by exaggerating the effect of intrinsic frequency in high-frequency list and reversing it in low-frequency list).

Connine et al.'s results did not support these predictions. They found a strong effect of extrinsic frequency for fast responses and a reduced effect for slow responses. Moreover, in the mixed lists, intrinsic word frequency influenced the identification responses, but this influence was equivalent for all ranges of response times. Connine et al. interpreted these results as evidence against an implementation of (intrinsic) frequency in resting activations, as in TRACE. They argued that in TRACE, frequency effects should be greater for longer response times because the system would have more time for interaction between lexical representations, which code lexical frequency, and phonemic representations, which phoneme identification is assumed to bear on. Connine et al. concluded that intrinsic word frequency does not function as an early source of information used to shape lexical hypotheses, as would be predicted if frequency were coded in resting activations. Rather, they argued that frequency affects postaccess decision processes.

Although the Connine et al. (1993) study provides intriguing support for frequency as a late response bias, the results are far from conclusive. First, the logic of the Connine et al. argument requires two controversial assumptions made by TRACE. The first is that there is feedback from lexical representations to phonemic representations. The second is that phonemic decisions are made using the same phonemic nodes that are used in lexical access. Connine et al.'s predictions would not hold for a model without feedback or for a model that posits separate decision nodes (cf. Norris, McQueen, &

Cutler, 2000). Furthermore, Connine et al.'s arguments are based on verbal predictions rather than actual simulations with TRACE. Thus, it is difficult to know under just what conditions the predictions hold. Making predictions without simulations is not straightforward without explicit assumptions about how to relate the time course of lexical-activation levels and fast and slow response latencies in phoneme identification. Finally, McQueen (1991) found larger lexical influences on fast phoneme-identification times than on slower identification times. This can be interpreted either as evidence against feedback between lexical and phonemic levels or as evidence that phoneme-identification latencies do not capture the time course of lexical influence on phoneme activation: By the time fast responses are generated, lexical feedback to phonemic nodes may have largely taken place; lexical influence on slow latencies may be masked by the influence of late decision biases specific to the task or activations reaching asymptotic levels. Regardless of how McQueen's results are interpreted, they suggest that using phoneme-identification times to track the time course of lexically based frequency effects is not straightforward.

The issues raised by the Connine et al. (1993) study highlight the importance of having precise information about the time course of activation for lexical competitors in order to identify where, in the recognition process, word frequency operates. Although researchers have developed an arsenal of useful experimental methodologies (see Grosjean & Frauenfelder, 1996; Lively, Pisoni, & Goldinger, 1994), it remains difficult to obtain data about lexical access in continuous speech that is fine-grained enough to favor one word-frequency account against the others. The presentation of increasingly longer word fragments to measure the degree of activation of high- and low-frequency competitors (Grosjean, 1980; Tyler, 1984; Zwitserlood, 1989) aims at providing a continuous measure of lexical activation. However, because it is quite remote from normal listening conditions, this technique may cause listeners to adopt strategies that might not reflect normal lexical access.

A growing number of researchers, building on work by Cooper (1974), have recently begun to use eye movements to explore questions about the time course of spoken-language comprehension (e.g., Altmann & Kamide, 1999; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Keysar, Barr, Balin, & Brauner, 2000; Tanenhaus & Spivey-Knowlton, 1996; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999). In the version of the eye-tracking paradigm introduced by Tanenhaus et al. (1995), participants follow spoken instructions to manipulate real or pictured objects displayed on a computer screen while their eye movements to the objects are monitored using a lightweight camera mounted on a head band. Eye movements to objects in the visual scene have been found to be closely time-locked to referring expressions in the unfolding speech stream.

Allopenna, Magnuson, and Tanenhaus (1998) explored the application of this technique to the study of spoken-word recognition. Participants were instructed to fixate a central cross and then followed a spoken instruction to move (using a computer mouse) one of four objects displayed on a computer screen (e.g., ‘‘Look at the cross. Pick up the beaker. Now put it above the square’’). Eye movements to each of the objects were recorded as the name of the referent object unfolded over time. On some crucial trials, the names of some of the distractor objects were phonologically similar to the name of the referent. For instance, the target picture *beaker* was presented with the picture of a competitor that overlapped with the target word at onset, *beetle* (henceforth, a *cohort* competitor, predicted to compete by the ‘‘Cohort’’ model, e.g., Marslen-Wilson & Welsh, 1978). The probability of fixating each object as the target word was heard was hypothesized to be closely linked to the activation of the lexical representation of this object (i.e., its name). The assumption providing the link between lexical activation and eye movements is that the activation of the name of a picture influences the probability that a subject will shift attention to that picture and thus make a saccadic eye movement to fixate it.

Allopenna et al. (1998) showed that eye movements generated early in the target word were equally likely to result in fixations to the cohort competitor (e.g., *beetle*) and to the referent (e.g., *beaker*) and were more likely to result in fixations to these pictures than to distractor controls that were phonologically unrelated to the target word (e.g., *carriage*). Furthermore, the fixations over time to the target, the cohort competitor, and a rhyme competitor (e.g., *speaker*) closely matched functions generated by the TRACE model of spoken-word recognition, given a simple implementation of the hypothesis linking activation levels in TRACE to fixation probabilities over time. The Allopenna et al. study suggests that the eye-tracking paradigm is a powerful tool for providing detailed time-course information about lexical access in continuous speech. The timing of frequency effects as revealed by fixations should be informative about whether word-frequency operates early in processing, when the input is still ambiguous among multiple lexical alternatives, or late in processing, after the input has converged on a single lexical candidate. Furthermore, the explicit linking hypothesis between lexical activation and observed fixations allows for quantitative comparisons of the goodness of fit between the observed and predicted fixations when word frequency is implemented in resting activations, in connection strengths, or as a bias in computing the activation of word-decision units.

Examination of frequency effects is also important for evaluating the general usefulness of the eye-tracking paradigm. In this paradigm, a small set of pictures is visually available to the listener. Participants could be adopting task-specific strategies that bypass ‘‘normal’’ language comprehension (e.g., preactivating the names of the visually present pictures). Thus, it is possible that the lexical candidates that enter the recognition process may be restricted to the visually present alternatives. Previous research has shown that well-documented frequency and neighborhood effects in word recognition can be dramatically reduced or even disappear in closed-set tests when all response alternatives are treated as equally probable (Pollack, Rubenstein, & Decker, 1959; Sommers, Kirk, & Pisoni, 1997). Thus, it remains to be established whether the eye-tracking paradigm is sensitive to characteristics of the lexicon that are not directly represented in the set of pictures displayed on a trial. The answer bears directly on the hypothesis linking lexical activation to eye movements and on the overall potential of the methodology for studying word recognition in continuous speech.

The present study had three goals. First, we asked whether the eye-tracking paradigm could capture subtle aspects of word processing such as word-frequency effects. Second, by analyzing the time course of the frequency effects on fixations, we evaluated two alternative accounts for the locus of frequency effects: the response-bias account which predicts late frequency effects, as suggested by Balota and Chumbley (1984, 1985) and Connine et al. (1993), and an account of frequency as part of the word-recognition system which predicts that frequency affects even the earliest moments of lexical activation. Finally, we implemented frequency in three ways (frequency in resting activation, frequency in connection strengths, and frequency as a postactivation bias) and compared the goodness of fit between the predicted fixations generated by these models and the data. These simulations evaluated whether these different models would yield different quantitative and/or qualitative predictions and whether the data would favor one over the others.

In Experiment 1, we presented participants with displays consisting of a referent along with two cohort competitors that varied in frequency and an unrelated distractor. For example, the referent *bench* was presented along with a high-frequency competitor, *bed;* a low-frequency competitor, *bell;* and a distractor, *lobster*. Participants were instructed to pick up the designated object by clicking on it with the computer mouse (e.g., ‘‘Pick up the bench’’). As the initial sounds of the target word were heard, the competitors were expected to be fixated more than the distractor as a consequence of their phonological similarity with the initial portion of the input. In addition, if fixations reflect lexical processing, more fixations to the high-frequency competitor than to the low-frequency competitor would be expected. Crucially, if lexical frequency operates on the lexical-access process (rather than as a possible response bias), the advantage for fixating the high-frequency competitor over the low-frequency competitor should be observed before the auditory input provides disambiguating information.

Finding a frequency effect on the fixations to visually present competitors would demonstrate that participants do not treat them as equally probable referents. However, finding such effect does not preclude the possibility that participants adopt some sort of verification strategy that does not reflect normal spoken-word processing. In particular, participants could name the visually present pictures to themselves before hearing the referent's name and store these names in short-term memory; frequency could bias the order in which the preactivated names are stored and/or retrieved. Experiment 2 evaluated whether an effect of frequency could be obtained when none of the distractors was phonologically similar to the target. We varied the frequency of the names of the referents (e.g., the high-frequency target *horse* vs the low-frequency target *horn*) and presented them along with three phonologically unrelated distractors. The referent picture was thus the only picture with a name matching the target word. If the probability of fixating a picture reflects activation of the lexical representation associated with this picture, fixations should be made to referent pictures with high-frequency names faster than to referent pictures with low-frequency names because high-frequency words are activated faster than low-frequency words. Accounting for such a frequency effect by a verification strategy due to previewing the pictures before any relevant information is heard is not tenable, as we discuss later.

EXPERIMENT 1

Method

Participants

Twenty native speakers of English were recruited at the University of Rochester and paid $7.50 for their participation.

Materials

Seventeen triplets were constructed. Each consisted of a target and two cohort competitors. Cohort competitors were chosen in such a way that they overlapped with the target's phonological form to the same extent, but differed in lexical frequency (as reported in Francis & Kučera, 1982). For example, one target was *bench,* with the high-frequency competitor *bed* (with a frequency of 139 per million) and the low-frequency competitor *bell* (with a frequency of 23 per million). For one triplet, *bandaid/bank/banjo,* the low-frequency competitor overlapped with the target word more than the high-frequency competitor did. The complete set of triplets is presented in Appendix A, set A. On average, the high-frequency competitor had a frequency of 138 per million; the low-frequency competitor, 10 per million; and the target, 14.5 per million. Each triplet was associated with a phonologically unrelated distractor (e.g., *lobster*). In addition to these 17 experimental displays, 23 filler displays were constructed. In order to prevent participants from developing expectations that pictures with phonologically similar names were likely to be targets, 8 of the filler trials contained three items that started with similar sounds and a phonologically unrelated item, which was the target. The 15 other filler trials were composed of four phonologically unrelated items; three of them were presented at the beginning of the session to familiarize participants with the task and the procedure.

The 160 pictures [(17+23) trials $\times$ 4 pictures] were all black-and-white line drawings. They were selected from the Snodgrass and Vanderwart (1980) picture set as well as from children's picture dictionaries. In order to ensure that the pictures associated with high- and low-frequency items were equally identifiable, we presented the pictures to 18 participants and asked them to write the name of the object represented. We thus collected names for the 34 cohort-item pictures. A correct response was an answer that exactly corresponded to the intended name or this name preceded or followed by a modifier. So, *shopping cart* for the intended *cart* was coded as a correct response, but *chest* for the intended *trunk* was not. The agreement between participants' responses and the intended names was 91.4% for the low-frequency cohorts and 90.1% for the high-frequency cohorts.

Despite similar name agreement for high- and low-frequency items, it was important to control for possible visual differences. Indeed, if the pictures associated with high-frequency competitors attracted more attention than the pictures associated with low-frequency competitors, more and longer fixations to the former compared to the latter would be found and mistakenly interpreted as a frequency effect. We thus needed, for each type of competitors, an estimate of fixation probability that was independent of their phonological similarity with the target word. We thus constructed another set of experimental trials and fillers (set B). The experimental trials consisted of the same items except for the target, which was phonologically unrelated to the cohorts or to the distractor (e.g., *mushroom* was one target, presented along with *bed, bell,* and *lobster,* and participants heard the instruction ''Pick up the mushroom''). These trials allowed us to compare the probability of fixating each cohort item when neither matched the acoustic information of the target word. These probabilities should be triggered by the visual characteristics of the pictures. In order to prevent strategies for finding a trial's target on the basis of phonological similarity between the pictured items, we constructed 13 filler trials where two items shared some similarity and one of them was the target. The materials for set B are presented in Appendix A. Ten participants were randomly assigned to each display set.[1] For each set, five random orders were created; approximately the same number of participants were assigned to each order.

The spoken instructions were recorded by a male native speaker of English in a soundproof room, sampling at 22,050 Hz with 16-bit resolution. Each instruction was then edited and some basic duration measurements were made. On average, the *pick up the* part of the instruction was 402 ms long in set A and 371 ms in set B; the target word was 498 ms long in set A and 497 ms in set B.

Procedure

Participants were seated at a comfortable distance from the computer screen. Participants' eye movements were monitored using an Applied Scientific Laboratories eye tracker. Two cameras mounted on a lightweight headband provided the input to the tracker. The eye camera provided an infrared image of the eye. The center of the pupil and the first Purkinje image (corneal reflection) were tracked to determine the position of the eye relative to the head. A scene camera was aligned with the participant's line of sight. A calibration procedure allowed software to superimpose crosshairs showing the point of gaze on a HI-8 videotape record of the scene camera. The scene camera sampled at a rate of 30 frames per second, and each frame was stamped with a time code. Auditory stimuli were played to the subject through headphones and simultaneously to the HI-8 VCR, providing an audio record of each trial. Two different computers were used to present the visual and the auditory stimuli on each trial, and the experimenter synchronized these two events by pressing both keyboards simultaneously. Note that timing measurements were made independently of the accuracy of this synchronization because speech and eye movements were assessed directly from the video recording. Any variability in this synchronization resulted only in slight variance in the delay between the presentation of the pictures and the onset of the spoken instruction.

[1] Each participant was actually exposed to both sets and the order of presentation of the sets was varied between participants; however, we only analyzed the trials for the first set presented.

The structure of each trial was as follows: First, a 5 × 5 grid with a centered cross appeared on the screen, and participants were instructed to look at the cross and to click on it. This allowed the experimenter to check that the calibration of the eye tracker was satisfactory. (Note that this instruction was given before the pictures were displayed, and participants were not instructed to fixate the cross at any other moment during the trial). Four line-drawing pictures and four colored geometric shapes appeared on specific cells of the grid. Participants were seated between 40 and 60 cm from the screen; each cell in the grid subtended 3° to 4° of visual angle, which is well within the resolution of the tracker (better than 1°). Approximately 500 ms after the pictures appeared, the spoken instruction started. The format of the instruction was constant across all trials: Participants were first asked to pick up one of the four pictures using the computer mouse (e.g., ''Pick up the bench.'') and then to move the picture above or below one of the four geometric shapes (''Put it above/below the circle/square/diamond/triangle.''). Once this was accomplished, the next trial began. The positions of the geometric shapes were fixed from one trial to the next. The position of each picture was randomized for each subject and each trial.

To minimize participants' prior exposure to the pictures, we departed from the procedure used in Allopenna et al. (1998) in two ways. First, the set of pictures was not shown to the participants before the experiment. Second, Allopenna et al. gave participants approximately 2 s to inspect the pictures after they had appeared on the screen and before being instructed to fixate the cross until the critical instruction began. The advantage of this procedure is that nearly all fixations begin on the cross. However, a disadvantage is that it provides participants with some time to inspect the pictures. With the current procedure, the delay between the presentation of the pictures and the spoken instruction was only 500 ms, making it less likely that participants would have time to implicitly name the pictures. Participants tended to make an eye movement to one of the pictures as soon as the pictures were displayed; therefore they could be fixating any of the four objects at the onset of the target word.

Data Collection and Coding

The data were collected from the videotape records using an editing VCR with frame-by-frame controls and synchronized video and audio channels. Coders used the crosshairs generated by the eye tracker to indicate where participants were looking at each video frame (30 per s) of the test trials. Fixations were coded on each trial from the onset of the target noun until the subject had moved the mouse cursor to the target picture. The onset of the target word on each trial was determined by monitoring the audio channel of the VCR frame by frame. Coders noted the onset of the instruction ''Pick up the . . .''; this time, plus the duration of the *Pick up the* instruction (independently measured with a speech-waveform editor), was identified as the onset of the target word.

For each subject and each trial, we established which of the four pictures or the cross was fixated at each time frame, beginning at the onset of the target word. The subject's gaze had to remain on the object for more than one frame to be counted as a fixation. If blinking occurred, fixation data was lost, typically for one to three frames. This time was attributed to the previous object being fixated, which was the best inference we could make about a participant's object of attention during a blink. Saccades from one fixation point to another were generally performed within one time frame; in the rare cases where it took more than one frame to reach a new object (at most two frames), the saccade time was also added to the fixation time on the previous fixation point.

Results and Discussion

Analysis of Set A Data

For two participants, one trial was missing because of technical failures. In order to give these participants' data the same weight as the data for the other participants, fixation values for these missing trials were estimated by using the participants' average proportions over the remaining trials. Figure 1 presents the proportions of fixations to the target, the averaged cohort competitors, and to the distractor in 33-ms time slices from 0 to 1000 ms after target onset. As is apparent in the figure, the fixation proportions for the target and competitors began diverging from the distractor shortly after 200 ms. The minimum latency to plan and launch a saccade is estimated to be between 150 and 180 ms in simple tasks (e.g., Fischer, 1992; Saslow, 1967). Moreover, saccadic eye movements are ballistic; once they are programmed, the fixation target is fixed. Therefore, a saccade initiated at 300 ms could only be influenced by acoustic information in the first 100 ms of a word. Thus, 200 ms after target onset is approximately the earliest point at which one expects to see fixations driven by acoustic information from the target word. A one-way ANOVA conducted on the mean proportion of fixations to the target picture, the high-frequency competitor, the low-frequency competitor, and the distractor, over the time window extending from 0 to 200 ms after target onset, showed no significant difference [$F1(3, 27) = 0.11$;

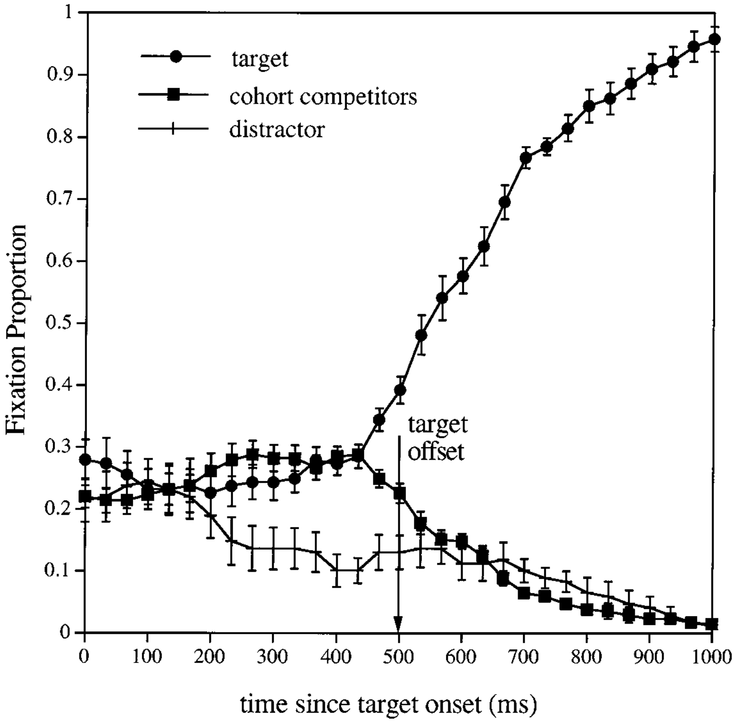

FIG. 1. Experiment 1: Fixation proportions over time for the target, the two averaged cohort competitors, and the distractor on the trials from set A. Bars indicate standard errors.

$F2(3, 48) = 0.14$]. The fixations to the averaged cohort competitors were significantly higher than the fixations to the distractor from about 200 ms until about 500 ms, when they merged, while the fixations to the target kept rising. A one-way ANOVA conducted on the mean fixation proportion to the target, averaged cohort competitors, and distractor, over a time window extending from 200 to 500 ms, revealed a significant effect of picture [$F1(2, 18) = 13.71$, $p < .0001$; $F2(2, 32) = 6.17$, $p < .01$]. Planned t test comparisons between the cohorts and the distractor fixations over the 200- to 500-ms time window revealed a significant difference [27.3% vs 12.8%, $t_1(9) = 4.14$, $p < .005$; $t_2(16) = 3.36$, $p < .005$]. Over most of this time window, the fixations to the target and the cohorts were similar, with no significant difference between the cohorts and the target fixation proportion between 200 and 433 ms. This suggests that between 200 and 500 ms, the cohort competitors were activated and competed with the target for recognition, until the acoustic information provided disambiguation. Note that this window is narrower that the one found in Allopenna et al. (1998), which extended from 300 to 700 ms. This difference can be accounted for by the fact that the mean overlap between the target and cohort competitor was greater in the Allopenna et al. study (3.38 phonemes) than in the present study (2.18 phonemes).

Having established the basic ‘‘cohort’’ effect and the time interval over which it was observed, we can now ask whether the high-frequency competitor was fixated more than the low-frequency competitor over this time interval. Figure 2 shows the proportion of fixations to the high- and low-frequency competitors, along with the proportions of fixations to the target and the distractor. As is apparent from the figure, the fixations to high- and low-frequency competitors started diverging at about 267 ms. When saccadic programming time and the ballistic nature of saccadic eye movements are taken into account, this means that frequency affected eye movements as early as the first 100 ms of the spoken word. The fixation proportion to the target picture and to the low-frequency competitor remained comparable until 467 ms after target onset (this result was expected because the frequencies of the target word and the low-frequency competitor were similar); after this point, the fixations to the target surpassed all other fixations. At about 467 ms, fixation proportion to both high- and low-frequency cohorts began dropping, while fixation proportion to the target began rising, presumably triggered by disambiguating acoustic information. The difference in fixations between high-frequency and low-frequency competitors extended until about 533 ms, although this difference started diminishing by 467 ms. A one-way ANOVA revealed a significant effect of picture (target, high-frequency competitor, low-frequency competitor, and distractor) over the 200- to 500-ms window [$F1(3, 27) = 8.85$, $p < .0001$; $F2(3, 48) = 5.13$, $p < .005$]. Over this time window, the high-frequency competitor was fixated more than the low-frequency competitor [31.5% vs 23.1%, $t_1(9) = 1.86$, $p < .05$; $t_2(16) =$

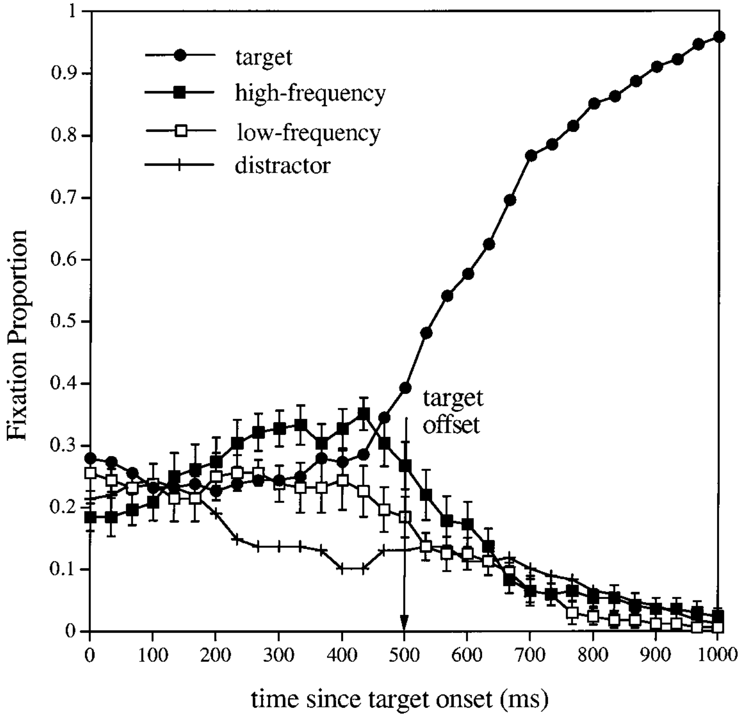

FIG. 2. Experiment 1: Fixation proportions over time for the target, the high-frequency cohort competitor, the low-frequency cohort competitor, and the distractor on trials from set A. Bars indicate standard errors.

2.10, $p < .05$]. These results indicate that fixations to each picture varied with the similarity between the phonological representation associated with the picture and the sensory input as well as its lexical frequency.

We also conducted a contingent analysis by selecting the trials on which participants were fixating either the distractor or the cross at the onset of the target word (47 of the 168 trials, 28%). If participants happened to be looking at the target or at one of the cohort competitors at the onset of the target word, they may have kept fixating this picture as the spoken input unfolded, as long as its name was consistent with the spoken input. This analysis ensures that any advantage observed for fixating the high-frequency cohort over the low-frequency cohort was not influenced by where the subject happened to be fixating at the onset of the target word. Figure 3 presents the fixation proportions for the target, each cohort competitor, and the distractor, from 0 to 1000 ms after target onset, on this subset of the data. As is apparent in the figure, the proportion of fixations to the high-frequency cohort was greater than the proportion of fixations to the low-frequency cohort, and the magnitude of the frequency effect was larger for this subset of data than for

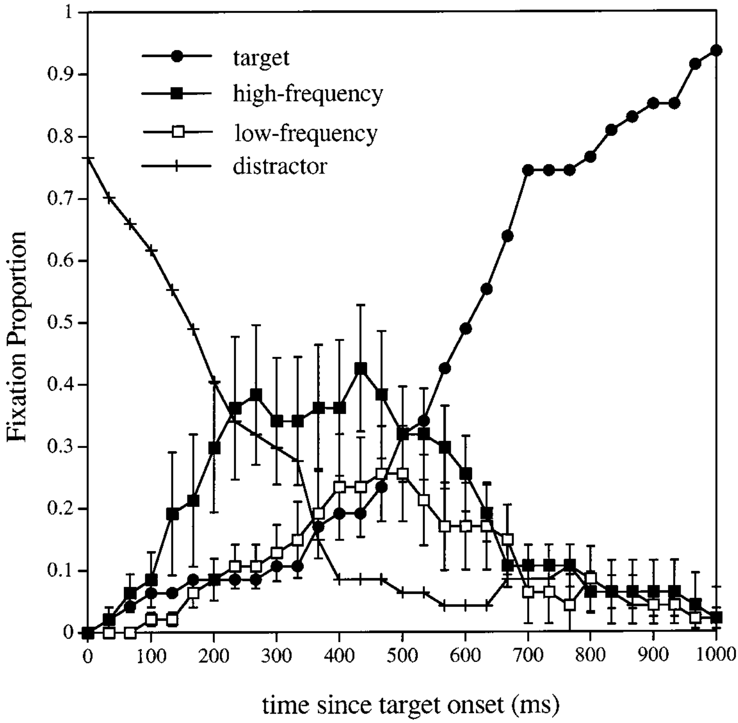

FIG. 3. Experiment 1: Fixation proportions over time for the target, the high-frequency cohort competitor, the low-frequency cohort competitor, and the distractor restricted to the trials from set A that started on the distractor or on the cross at target onset. Bars indicate standard errors.

the full data. Variability (as indicated by the error bars) also increased because of the small number of observations included in the subset. Comparisons between the cohorts over the 200- to 500-ms window revealed a significant difference by subjects [$t_1(9) = 2.31$, $p < .05$], but only a marginal difference by items [$t_2(16) = 1.37$, $p = .09$]. The marginal result in the item analysis is most likely due to the relatively few observations that remained for each item after the trials with initial fixations to the cohort and target pictures were removed. In general, the contingent analysis confirms that the high-frequency competitor was fixated more than the low-frequency competitor.

Thus far, we have presented evidence that frequency influences fixations very early in the processing of the target word, as early as 267 ms after target onset. We conducted additional analyses to firmly establish that frequency affects lexical activation when the input is still ambiguous between several candidates and before a single candidate could be selected based on bottom-up input. Three phonetically trained native speakers of American English were instructed to determine the point, in the target word, where the signal could unambiguously distinguish the target from its competitors (e.g., where, in the word /bɛntʃ/, they could be certain that *bench*, and not *bed* or *bell*, was heard). Using a speech editor, they selected increasingly larger fragments of the target word until they reached a point where the target could be identified and the competitors rejected. The names of the targets and competitors were provided beforehand. This procedure provided, for each of the 17 items, an estimate of the uniqueness point (with respect to the other alternatives present on the display). The uniqueness-point estimates were quite similar across judges and correlated highly (mean $r = .83$); they were thus averaged. The mean uniqueness point was 237 ms (ranging from 128 to 352 ms). Uniqueness-point estimates allowed us to test whether frequency effect occurred as soon as candidates were activated (i.e., before the uniqueness point), as predicted by a frequency mechanism operating very early, or when only one candidate matched the signal (i.e., after the uniqueness point), as predicted by a late response-bias account.

For each of the 17 items, we computed the mean proportion of fixation to the high-frequency and low-frequency competitors over two time windows: (1) over the time window extending from the onset of the target word plus 200 ms to the uniqueness point plus 200 ms—the window where fixations could plausibly be affected by signal information from target onset up until the uniqueness point; and (2) over the time window extending from the uniqueness point plus 200 ms to the target offset plus 200 ms—the window where fixations could have been influenced from signal information coming at or after the uniqueness point. On average, the high-frequency competitor was more likely to be fixated than the low-frequency competitor, for each time window (32% vs 24% for the preuniqueness point window and 19% vs 13% for the postuniqueness point window). A two-way (window $\times$ competitor frequency) ANOVA revealed a significant effect of competitor frequency [$F(1, 16) = 4.7$, $p < .05$] and of window [$F(1, 16) = 29.5$, $p < .0001$] and no significant interaction ($F < 1$). This demonstrates that the frequency effect was not limited to the postuniqueness point window, as predicted by a late response bias account of frequency, but was already present in the preuniqueness point time window.

Analysis of Set B Data

In this set, both cohort pictures were presented along with a phonologically unrelated target (e.g., the cohort pictures *bed* and *bell* were presented along with the target *mushroom*). Fixations to these pictures as the target word was heard should have been triggered by the visual characteristics of the pictures only, since their names were phonologically dissimilar to the target. For two participants, a few trials were missing because of technical failures (one for one participant and two for the other participant). Figure 4 presents the fixation proportion over time for the target, the high- and low-frequency

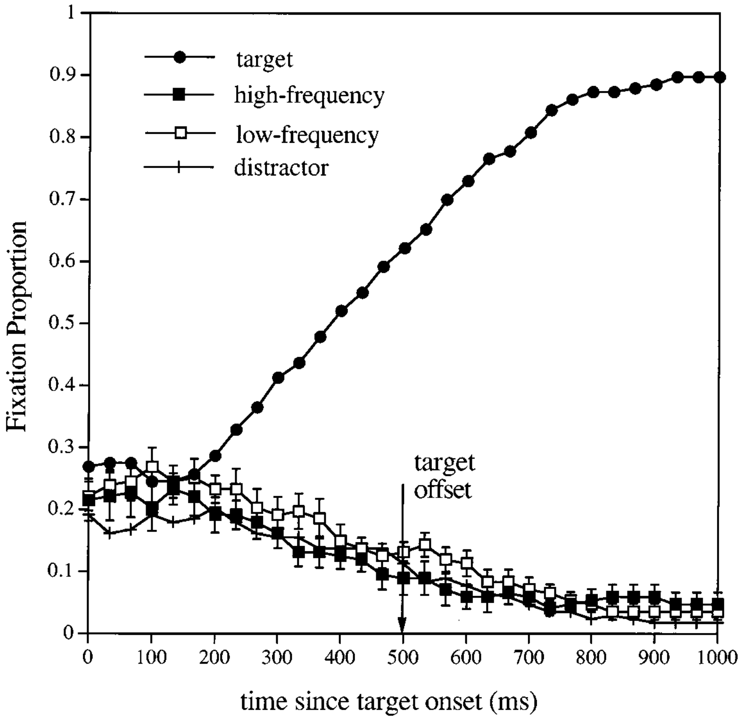

FIG. 4. Experiment 1: Fixation proportions over time for the target, the high-frequency cohort item, the low-frequency cohort item, and the distractor on the trials from set B. Bars indicate standard errors.

cohort items, and for the distractor. Fixations to both cohort items and to the distractor steadily decreased while fixations to the target increased as the target word was heard. Crucially, in the 200- to 500-ms window, the fixation proportion for the high-frequency cohort item did not surpass that for the low-frequency cohort item—in fact, the low-frequency cohort item showed a small advantage over its high-frequency counterpart, although the difference was not reliable [$t_1(9) = 1.22$, $p > .10$; $t_2(16) = 1.19$, $p > .10$). Thus, the analysis of the set B trials indicated that the visual characteristics of the pictures cannot account for the advantage for fixating high-frequency cohort items observed in set A. The effect can thus be attributed to the influence of lexical frequency on cohort activation.

Experiment 1 revealed three central results. First, it confirmed that the fixations to targets and competitors are tightly time-locked to the unfolding acoustic input, as demonstrated in Allopenna et al. (1998). Recall that in our study (as opposed to that of Allopenna et al.), participants had no exposure to the pictures prior to the experiment and quite limited exposure (500 ms) to the four pictures prior to the spoken instruction during each trial. Thus, the time locking is not a consequence of extensive exposure. Second, the results showed that the probability of fixating a competitor that shares phonological similarity with the target word are further influenced by the competitor's lexical frequency in the language. Even though the display limited potential responses to four alternatives (i.e., four pictures), an effect of word frequency was still observed (at least under conditions where the exposure to the pictures was limited). Finally, the timing of the frequency effect sheds light on the locus of frequency effects in word processing. The fixations to the high-frequency cohort competitor started diverging from the fixations to the low-frequency cohort competitor as early as 267 ms after target-word onset, reflecting fixations that were programmed within the first 100 ms of the beginning of the word, and while the input was still ambiguous among multiple lexical alternatives (i.e., prior to the uniqueness point). We thus see an effect of word frequency as early as it can be observed. This is incompatible with a view of word frequency operating as a late response bias after lexical access is complete.

Simulations

An important advantage of the eye-tracking paradigm is that hypothetical lexical activations can be mapped onto fixation proportions using an explicit, well-motivated linking hypothesis (Allopenna et al., 1998), allowing for quantitative tests of alternative models. In order to evaluate the various accounts of frequency and their fit to our data, we implemented frequency in the TRACE model (McClelland & Elman, 1986). We chose TRACE because it is a well specified and implemented model of spoken-word recognition and is publicly available. It accounts for a wide variety of phenomena (McClelland & Elman, 1986), including the time course of fixation probabilities in tasks like ours (Allopenna et al., 1998). TRACE shares several important basic properties with other current models of spoken-word recognition. Words are activated incrementally as an input sequence unfolds, proportionally to their similarity with the input, and activated words compete for recognition.

TRACE is an interactive activation network composed of three levels. Input arrives in the form of idealized, pseudospectral representations of features over time. The featural nodes have excitatory connections to appropriate phoneme nodes. Phoneme nodes in turn excite appropriate lexical nodes. Competition is implemented via lateral inhibitory connections within the phoneme and lexical levels. In addition, lexical nodes feed back via excitatory connections to the phonemes they are connected to. (Feedback connections also exist from phonemes to features, but the weights of these connections were set to 0 in McClelland & Elman, 1986, as well as in our simulations.) Input arrives in featural slices, with phonemes having durations of 11 slices. Because phonemes are centered 6 slices apart, input slices corresponding to a phoneme overlap with slices corresponding to the preceding and following phonemes. Activation spreads forward through the network from feature to phoneme to lexical nodes, after each input slice (with a 1-slice lag for word-to-phoneme feedback).

We compared three different implementations of frequency: frequency operating on resting-activation levels, frequency operating on connection weights, and frequency applied in a postactivation decision rule, as suggested by Luce and colleagues in the NAM (Goldinger et al., 1989; Luce, 1986; Luce & Pisoni, 1998). In each of these simulations, frequency is viewed as a central component of lexical access; that is, frequency plays a role as soon as lexical candidates become active (this is described in more detail below). The simulations were compared to simulations that did not incorporate frequency. These simulations had three goals. The first goal was to determine whether simulations with frequency would provide a better account for the data than simulations without frequency. Note that achieving better fits with frequency as a factor is not trivial because (a) phonological similarity accounts for most of the variance and (b) we are not using frequency as a free parameter; rather, we are using theoretically motivated hypotheses about how frequency affects activations to generate activation-driven predictions. The second goal was to use explicit implementations to determine what different predictions, if any, competing frequency accounts make about the time course of lexical activation. If significant differences arose, the third goal was to determine which frequency model best fits the data.

We used the publicly available implementation of TRACE.[2] We augmented the 235-word lexicon included with the distribution with approximations to all of the target and cohort stimuli (but not the distractors) from Experiments 1 and 2.[3] (Distractors were chosen from unrelated items in the lexicon.) We added 51 words for Experiment 1 and 17 words for Experiment 2, resulting in a 303-word lexicon.

[2] This is available from Jeff Elman at the UCSD Center for Research in Language via anonymous ftp at ftp://crl.ucsd.edu/pub/neuralnets/. Note that the parameter set included in this implementation was changed to match those reported by McClelland and Elman (1986)—with the exception of the frequency-scale parameter, which we varied in the simulations reported here. These parameters are reported in Appendix B.

[3] Because TRACE only incorporates a subset of English segments (the vowels /ɒ, u, i, ʌ/ and the consonants /b, p, d, t, g, k, s, ʃ, ɹ, d, l/), many of our stimuli could only be approximately transcribed. We tried to choose similar segments for substitutions without overusing any segment. We made no attempt to use transcription choices as free parameters. We transcribed the words only once and did not tweak our transcriptions to drive the performance of the model, with three exceptions: /tʌ/, /kʌ/, and /bʌl/ (original TRACE transcription for *toe, cow,* and *bell,* respectively). Because many words contain these syllables, these items did not become the most active items given appropriate input. We retranscribed the first two as /tʌʌ/ and /kʌʌ/, which approximates the fact that word-final vowels are lengthened in English, and allowed them to become the most activated words in response to appropriate input. We changed the vowel used for *bell* and *bed* to /i/ (transcribed as /bil/ and /bid/), which placed those items in a sparser neighborhood. Our transcriptions are listed in Appendices A and C.

Frequency in resting levels. The approach usually taken in interactive activation models is to make resting levels proportional to frequency (McClelland & Rumelhart, 1981). On this approach, each unit has a functional bias associated with it, such that high-frequency items have a head start in the form of higher activation in the absence of bottom-up input and their activations decay slightly more slowly. The TRACE implementation includes a method for incorporating frequency, but so far as we know, this aspect of the model has not been explored previously. In all the published reports we are aware of, frequency was turned off by setting the frequency-scale parameter to zero. However, a nonzero frequency value can be specified for each lexical item, which affects its resting-activation level. Resting level is determined by Eq. (1) as follows:

$$r_i = R + s[\log_{10}(c + f_i)] + p_i, \tag{1}$$

where r_i is the resting activation for unit i, R is the default resting activation, s is the frequency-scaling constant; f_i is the frequency of item i (number of occurrences per million words reported by Francis & Kučera, 1982; items not included in that corpus were given frequencies of 1), c is a constant that prevents taking the log of 0 or 1 and can also decrease the initial slope of the log transform (we set c to 1.0 for this method, so there was little effect on the slope), and p_i is a top-down factor (intended, for instance, for semantic-priming simulations) which was set to 0 for all our simulations.

Our initial explorations of this frequency implementation revealed two main constraints. First, if resting values are greater than 0, the activations of lower frequency items will be driven toward the minimum possible activation value by higher frequency nodes, even in the absence of bottom-up information. This is because units with positive activation can inhibit other units within the same layer. Therefore, in order for frequency biases to have a stable effect, even items with the highest frequencies must not have resting levels greater than 0. This means that the default resting level R has to be set sufficiently low that the frequency-scaling constant s can be set high enough to achieve a wide range of resting levels less than 0. We did two things to accomplish this. First, we set a ceiling for frequency values: Items with a frequency greater than 1000 were treated as having a frequency of 1000 (only eight items in the TRACE lexicon were affected, none of which were among our stimuli). Second, we set the default resting level R to -0.3. (We changed the TRACE implementation to report activations less than 0; note that our results could not be replicated without making this change.) This combination allowed a maximum value of 0.1 for the frequency-scaling constant s, which would give the highest frequency items resting levels of 0. We determined that a value of .06 gave the best balance of fits to the data from Experiments 1 and 2 with this method.

We converted activations to response probabilities using the Luce choice rule (R. D. Luce, 1959).[4] Activations, *a*, are converted to response strengths, *S*, as shown in Eq. (2), where *k* is a constant that determines the amount of separation between strengths. We set *k* to 7 for all of the simulations we report, as this was the fixed value that Allopenna et al. (1998) used to fit similar data. The exponential transformation in Eq. (2) ensures that no values are negative and amplifies large activation values. Response probabilities for each item, $P(R_i)$, are simply normalized response strengths, as shown in Eq. (3) as follows:

$$S_i = e^{ka_i} \tag{2}$$

$$P(R_i) = \frac{S_i}{\sum_{j=1}^{n} S_j}. \tag{3}$$

Note that the conversion to response probabilities embodies a simple assumption about the role of visual information in our task. Activations are generated in a bottom-up fashion on the basis of phonetic input. All lexical items enter into the activation and competition process. However, since only four responses are possible given the visual display, only the four displayed items enter into the decision rule. We refer to this method as $\mathrm{RDL}_{\mathrm{REST}}$ because it uses the R. D. Luce choice rule to evaluate activations with frequency instantiated in resting levels. We refer to simulations without frequency as the $RDL_{\mathrm{NO\ FRQ}}$ method.

Allopenna et al. (1998) used an additional step that scaled response probabilities to be proportional to the total amount of activation among the four

[4] Determining activations from TRACE is not a trivial process. Word units in TRACE function as templates. For a word unit to become highly active, it must be well aligned with phonemic (and featural) inputs. TRACE avoids the alignment problem by aligning a copy of each word unit every three input slices. Given input, TRACE reports the activity of copies of each word unit aligned at different slices. The experimenter must decide how to decode the word-unit patterns of activation. The method we used was to determine which copy of a word unit reached the highest activation and then use the activation of that unit over all input cycles as the activation of that word. We assume McClelland and Elman (1986) used a similar method when they examined activations for their simulations. They reported activations or response probabilities based on units aligned with a particular slice of the input, although they did not explicitly discuss how they chose that slice (see Frauenfelder & Peeters, 1998, for the description of a similar method). This procedure is problematic because it cannot be implemented in an incremental fashion; it requires an omniscient observer to compare peak activations after processing is finished. Incremental methods are possible. Each lexical item could have an associated decision node that would either summate the responses of all copies of the word template at all slices or report the activation of the most active word template at each slice. For the purposes of the current article, we use the simple method we have described and leave this issue open for future research.

possible visible targets at each time slice. This was necessary because in the Luce choice rule, given equivalent activations, the minimum response probability for any item is $1/n$, where n is the number of possible responses. Allopenna et al. instructed participants to fixate a central fixation cross immediately before the critical instruction was heard, with the result that the fixation proportions at the onset of target words were almost always 0 for all objects (since participants were fixating the central cross). For the current experiment, we did not give participants any explicit instructions regarding fixations. Participants were thus about equally likely to be fixating any picture in the display at the onset of target words. Thus, the basic choice rule, which assigns response probabilities of $1/n$ in the absence of bottom-up input, maps perfectly onto our task.

Frequency in connection weights. Another way to incorporate frequency in interactive activation models, advocated by MacKay (1982, 1987), is to make the weights (or connection strengths) associated with lexical units proportional to their frequency. Although implementations of frequency in resting levels or in connection weights are very similar to each other (e.g., Dell, 1990), our simulations show that the two implementations differ in one crucial respect. In order to implement frequency in the connection weights, we scaled phoneme-to-word input according to lexical frequency according to Eq. (4) as follows:

$$a'_{pi} = a_{pi}[1 + (a_{pi}s[\log_{10}(c + f_i)])], \tag{4}$$

where a_{pi} is the activation to lexical unit i from phoneme p. We found that a value of .13 for s achieved the best balance of fits for the data from Experiments 1 and 2, and c was set to 1.0. Note that with frequency implemented in connection weights, the scaling factor s is not constrained in the same way as for frequency implemented in resting levels, since units remain at the same default resting level until they receive bottom-up input. Activations were converted to response probabilities using the Luce choice rule as in the resting-level method. We refer to this method as RDL_{WT}.

Frequency in the decision rule. The final method used to incorporate frequency was based on the principles of the NAM. In this model, units corresponding to acoustic-phonetic patterns are activated on the basis of bottom-up input. These pattern units connect to decision units, which correspond to different lexical items. Decision units compute the *Stimulus Word Probability* (SWP) for their corresponding lexical item on the basis of the bottom-up pattern input (that is, the probability that the lexical item is being heard given the current input). They also have access to higher level lexical information, such as word frequency, and have access to the summed SWPs for all other lexical items. Based on these input sources, decision units compute the probability of identifying their corresponding word according to Eq. (5) as follows:

$$P(ID) = \frac{SWP_i f_i}{\sum_{j=1}^{n} SWP_j f_j}, \quad (5)$$

where i is the lexical item corresponding to the decision unit, f_i its frequency, and the denominator is the summed frequency-weighted SWPs for all items, including i.[5] Note the similarity between Eqs. (3) and (5); Eq. (5) would be a frequency-weighted variant of the R. D. Luce (1959) choice rule if TRACE response strengths were used as SWP estimates. While Luce (1986) and Luce and Pisoni (1998) used segmental confusion matrices to estimate SWPs, they also suggested that a TRACE-like interactive activation system could form the front-end to the NAM, such that SWPs would correspond to activations. A key difference in such an implementation compared to the standard TRACE model is the locus of frequency effects. When frequency is incorporated into the activation component of an interactive activation model like TRACE, its effects percolate throughout the system. In contrast, if frequency only comes into play at a postactivation decision stage, its effects will be confined to this decision stage, and Goldinger et al. (1989) predicted that these two implementations ought to have very different results. The ramifications of the two implementations are complex enough that simulations are called for to test in what ways they are different. In order to implement a decision rule that incorporates frequency independently of activations, we simply used TRACE activations (without frequency) as estimates of SWPs. This is equivalent to frequency-weighting response strengths, as in Eq. (6) as follows:

$$S_i = SWP_i = e^{ka_i} [\log_{10}(c + f_i)], \quad (6)$$

where a_i is the activation of the lexical unit for word i, f_i is its frequency, and c is a constant. Unlike the other simulation methods, we found that fits were substantially improved by manipulating the parameter c. A value of 15 was used for c for both experiments (making c larger improves the fit for Experiment 2, but decreases the fit for Experiment 1). Alternatively, one could use a frequency scaling parameter to control the influence of frequency in Eq. (6). However, scaling f by values greater than 1 has similar effects as increasing c, and values less than 1 have the undesirable effect of making the log-transform more linear. Response strengths were converted to response probabilities using Eq. (3). We refer to this simulation method as the

[5] Note that Luce and his colleagues usually separate the SWP of item i from the SWPs of the phonological neighbors of this item, which they denote as NWPs (neighbor-word probabilities). Without an *a priori* way of dividing the lexicon into neighbors and nonneighbors, it is simpler to refer to the SWPs of every word given the current input (most will approach 0).

RDL_{POST} method because it is the Luce choice rule with frequency applied at the ''postactivation'' stage.

Note that this implementation is somewhat inconsistent with the way the NAM has been described by Luce and colleagues. According to Goldinger et al. (1989), in the NAM, ''the effects of frequency are not realized until the selection phase of the recognition process [. . .] frequency is assumed to exert its influences *after* initial activation but *before* lexical access occurs'' (p. 504). Thus, frequency should not come into play during the initial processing, whereas in our implementation, frequency comes into play immediately, given nonzero activation. We did not implement a staged application of frequency of this sort. However, we can infer from the results of the $RDL_{NO\ FRQ}$ (without frequency) and RDL_{POST} simulations what the results of such an implementation would be: Early on, a system with a late-acting frequency mechanism would resemble TRACE without frequency, i.e., $RDL_{NO\ FRQ}$. Once some sort of threshold were reached (e.g., one or more items crossed an activation threshold), frequency would come into play, and the results would resemble RDL_{POST} from that point onward.

Simulation results. We conducted our simulations by presenting each target word to TRACE, one at a time. All lexical items were allowed to compete. Figure 5 shows the raw TRACE activations averaged over all the stimulus items. In the top panel, activations are shown with no frequency influence (activation for the high-frequency and the low-frequency competitors are thus identical); in the middle panel, frequency is incorporated into resting activations; and in the bottom panel, frequency is incorporated into connection weights.[6]

We then computed response probabilities for each of the four methods ($RDL_{NO\ FRQ}$, RDL_{REST}, RDL_{WT}, and RDL_{POST}). A crucial issue was how to map cycles onto real time. We used stimulus length to equate simulation time and real time. Our stimuli for Experiment 1 had a mean duration of 498 ms. Our stimuli, as represented in TRACE, had a mean duration of 40.1 cycles. If we equate stimulus time between our natural speech and TRACE representations, each cycle would be equivalent to 12.4 ms. Given that we sampled eye movements at 30 Hz (i.e., every 33.33 ms), we aligned cycles to milliseconds by linearly interpolating 11 intermediate steps between cycles (making each new cycle equivalent to 1.03 ms) and then downsampling to every 32nd point. Thus, each downsampled cycle corresponded to 33.12 ms. Allopenna et al. (1998) were able to equate 1 cycle to 11 ms, and, therefore, 3 cycles to one video frame, because their talker had a slightly faster speech rate. Note that it is quite remarkable that the temporal dynamics of TRACE allow such a principled method for equating simulation time and real time.

[6] Activations vary from −100 to 100 because TRACE multiplies the output values by 100. Before computing response probabilities, we multiplied the activations by .01 to convert them back to a scale varying from −1 to 1.

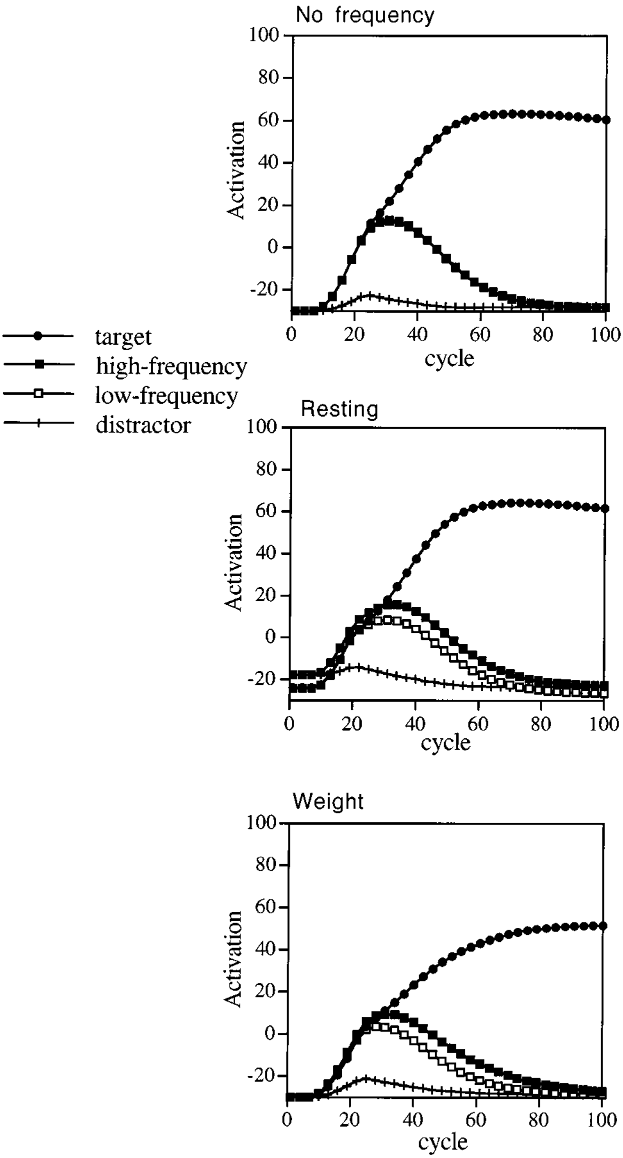

FIG. 5. Experiment 1 simulations: TRACE activations over time for the target, the high- and low-frequency competitors, and the distractor when frequency is turned off (top), when frequency is coded in resting activations (middle), and when frequency is coded in connection weights (bottom).

Another important issue is how to align simulation points and data points. Allopenna et al. (1998) added six extra frames of no activation to the beginnings of their simulations, which they equated with the time it takes to plan and launch an eye movement. We found that our various simulation methods required slightly different alignments for optimal fits to the data, so we treated alignment as a free variable (two extra frames of no activation were added for $\mathrm{RDL}_{\mathrm{POST}}$, three for $\mathrm{RDL}_{\mathrm{REST}}$ and $\mathrm{RDL}_{\mathrm{NO\ FRQ}}$, and four for $\mathrm{RDL}_{\mathrm{WT}}$; however, this only gave an average improvement of .07 in r^2 compared to direct alignment).

In Fig. 6, activations have been transformed into response probabilities for each of the four simulation methods ($\mathrm{RDL}_{\mathrm{NO\ FRQ}}$, $\mathrm{RDL}_{\mathrm{REST}}$, $\mathrm{RDL}_{\mathrm{WT}}$, and $\mathrm{RDL}_{\mathrm{POST}}$).

Table 1 shows the root mean squared (RMS) error between the data and simulation methods computed over two different windows: all data points from 0 to 1000 ms and from 200 to 500 ms (the region where reliable cohort effects were observed in the data). RMS error gives a measure of the absolute fit. The means for each item are shown to help interpret the RMS values. (Since RMS is an absolute measure, an RMS of 0.1 could be quite good for data with a mean near 1.0, but could be quite poor for data with a mean near 0.) Table 2 shows the corresponding r^2 values. In addition, RMS and r^2 values computed on the difference between high- and low-frequency cohorts are reported in each table. We included this measure because all of the methods yielded relatively small RMS and large r^2 values, suggesting a good fit with the data. For RMS, this was because the frequency-based differences we observed were relatively subtle (even the no-frequency method yielded RMS values comparable to the other methods for the high- and low-frequency items). For r^2, this was because the differences from time step to time step are important. To achieve a high r^2 value, the simulated data points must show decreases and increases proportional to those found in the data. Taking a difference score allows us to make a stronger test: Do the simulated data points also capture the differences between high- and low-frequency competitors? We thus concentrate on these values when comparing the different simulations.

As is apparent in Tables 1 and 2, the $\mathrm{RDL}_{\mathrm{NO\ FRQ}}$ method did not fit the data as well as the other methods, especially in terms of the difference between high- and low-frequency competitors in the 200- to 500-ms region. The RMS is about twice as great as that for the other methods, and this implementation provides no correlation with the data over time. All the other methods fit the data well. The differences in fixation probabilities between the high- and low-frequency competitors are shown in Fig. 7. As can be seen in the figure, all three methods incorporating frequency provide similar predictions. The main difference is in the initial difference predicted by the $\mathrm{RDL}_{\mathrm{WT}}$ method. With this method, frequency effects are proportional to activation in the system. Prior to substantial bottom-up input, no frequency dif-

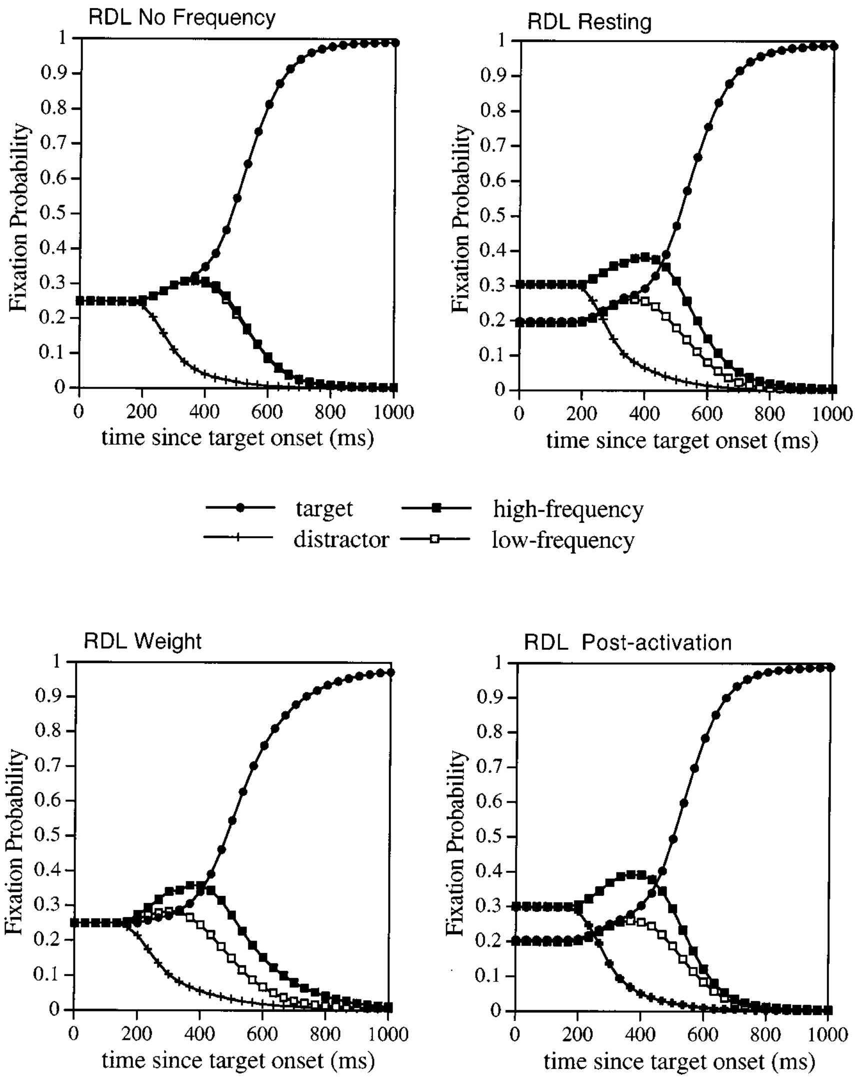

FIG. 6. Experiment 1 simulations: Fixation probabilities over time for the target, the high- and low-frequency competitors, and the distractor for each of the four frequency implementations (see text).

ference is predicted. As more input arrives, the frequency difference gradually increases. The other two frequency methods, RDL_{REST} and RDL_{POST}, predict a frequency bias, even with little bottom-up activation. Thus, the RDL_{WT} method is able to capture one significant feature of the data that the others cannot: a frequency effect with a gradual onset, modulated by bottom-up activity. It is in this respect that the resting-level and connection-strength

TABLE 1
RMS Measures of Model Fits to Experiment 1 for the Target, the High-Frequency (HF) and the Low-Frequency (LF) Cohort Competitors, and the Distractor, on the 0- to 1000-ms (All) and the 200- to 500-ms (Section) Windows

Mean	Item	$RDL_{NO\ FRQ}$	RDL_{REST}	RDL_{WT}	RDL_{POST}
.514	Target (All)	.117	.093	.090	.103
.278	Target (Section)	.081	.038	.075	.043
.182	HF Cohort (All)	.044	.053	.027	.055
.310	HF Cohort (Section)	.037	.040	.022	.048
.148	LF Cohort (All)	.038	.033	.031	.033
.232	LF Cohort (Section)	.053	.031	.029	.028
.125	Distractor (All)	.070	.076	.062	.076
.135	Distractor (Section)	.069	.072	.057	.074
.049	HF–LF (All)	.056	.069	.037	.069
.080	HF–LF (Section)	.086	.044	.029	.048

Note. Means for the data are shown in the leftmost column to guide interpretation of the RMS values.

implementations are not simple variants of each other; only the connection-weight implementation gives rise to a gradual (although still immediate and fast) onset of frequency.

All of the implementations that we have evaluated thus far assume that frequency applies early in the recognition process. However, one might defend a model in which frequency comes into play when the system is preparing to generate a response, that is, after lexical access and prior to generating an eye movement. This would be consistent with the late-acting response bias argued for by Connine et al. (1993). The argument would be that even when activation levels are low, there is still a small probability that activation

TABLE 2
r^2 Measures of Model Fits to Experiment 1 for the Target, the High-Frequency (HF) and the Low-Frequency (LF) Cohort Competitors, and the Distractor, on the 0- to 1000-ms (All) and the 200- to 500-ms (Section) Windows

Item	$RDL_{NO\ FRQ}$	RDL_{REST}	RDL_{WT}	RDL_{POST}
Target (All)	.955	.969	.965	.969
Target (Section)	.969	.963	.969	.968
HF Cohort (All)	.909	.931	.961	.931
HF Cohort (Section)	.606	.590	.638	.631
LF Cohort (All)	.942	.923	.940	.923
LF Cohort (Section)	.263	.091	.759	.090
Distractor (All)	.754	.761	.795	.756
Distractor (Section)	.672	.652	.691	.667
HF–LF (All)	.015	.106	.491	.139
HF–LF (Section)	.008	.342	.550	.475

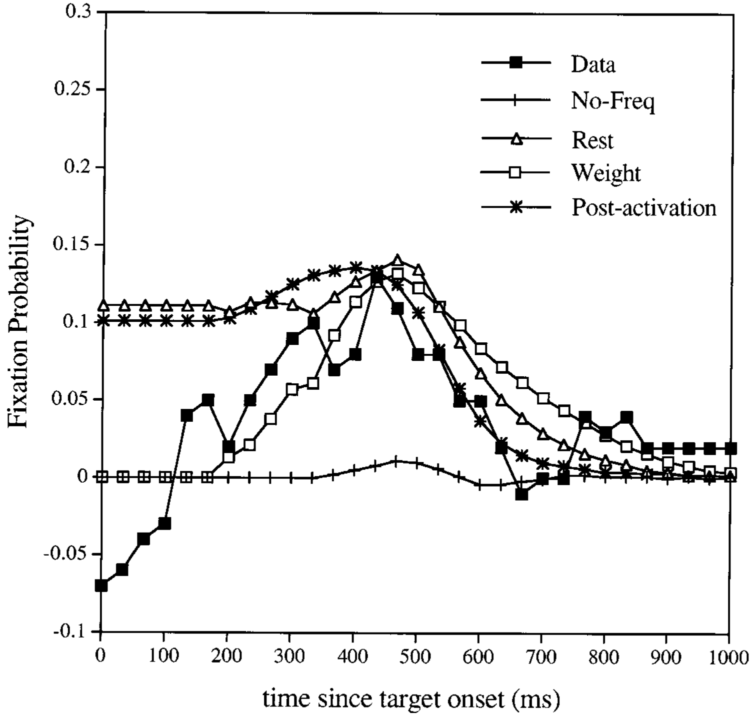

FIG. 7. Experiment 1 simulations: Fixation-probability differences between the high- and the low-frequency competitors over time for the data and each of the four frequency implementations (see text).

of a lexical alternative will cross a response threshold. Frequency would then operate as a constant bias and apply to lexical candidates that cross threshold. The RDL_{POST} method could be easily adapted to do something like this, e.g., by only allowing frequency to weight response strengths after a certain number of cycles of processing or when activations surpass a threshold. The result would resemble pasting the $RDL_{NO\ FRQ}$ and RDL_{POST} results together: Prior to the point at which frequency applied, this ''late'' method would look exactly like the $RDL_{NO\ FRQ}$ method. From the point at which frequency applied, it would look exactly like RDL_{POST}, with a sudden jump between the two patterns. However, there is no motivation to implement such a mechanism. Because frequency has a gradual and early onset in our data, there is no basis to prefer a late-acting, sudden-onset frequency bias. Moreover, modifying the postaccess model to assume that lexical access can take place as early as would be necessary to accommodate our data would clearly violate the spirit of the postaccess hypothesis: The threshold for generating a response after lexical access would have to be so low as to be indistinguishable from a model without a threshold.

Summary and Conclusions

Experiment 1 demonstrated that fixations to objects in the visual set that share phonological similarity with the sound pattern of the target word are affected by the frequency with which the names of the objects occur in the language. When both high-frequency and low-frequency cohort competitors were present, participants were more likely to fixate the high-frequency competitor than the low-frequency competitor. Furthermore, the difference in fixations between high- and low-frequency competitors was observed very early, when the acoustic signal was still lexically ambiguous. The results are consistent with predictions from models in which frequency affects lexical access as soon as lexical items are activated by bottom-up input. The TRACE model accounted for a large proportion of the variance in the time course of fixation probabilities, regardless of whether frequency was implemented in resting activation levels, connections weights, or as a bias applied continuously to (but independently of) activations (although the connection weight method yielded the best quantitative and qualitative fits).

EXPERIMENT 2

Experiment 2 was conducted to determine whether it was possible to observe an effect of target frequency when none of the other pictures had names that were phonologically related to the name of the referent. Demonstrating a frequency effect on target fixations, with no competitor pictures with names phonologically similar to the target's in the display, is important because it would provide a replication of the frequency effect using a different measure and would allow us to further evaluate the mechanism by which frequency operates during lexical processing. Moreover, finding such an effect on target fixations would be inconsistent with the possibility that participants adopt a verification strategy (i.e., naming the visually present objects prior to the spoken instruction and then matching the sound pattern of the target word with the preactivated names). This verification strategy could account for the frequency difference on fixations to cohort competitors in Experiment 1 by assuming that frequency biases the storage and/or retrieval of the preactivated names. However, because no frequency effect was found on fixations to the competitors in the set B data (when the target name did not overlap with the cohorts' names), one would also have to assume that the frequency bias in the verification strategy only applies to names that match the target word. In Experiment 2, a referent picture was presented along with three pictures with phonologically unrelated names. The name of the referent was either low-frequency or high-frequency. Assuming that high-frequency words are activated faster than low-frequency words, and that fixations are influenced by lexical activations, we predicted that participants would fixate the high-frequency referent picture (before clicking on it with the mouse) faster than the low-frequency referent picture. A verification strategy would not predict such a difference; both high-frequency and low-frequency referent pictures should be fixated equally fast because in both conditions, the target picture is the only visually present picture whose name matches the target word.

Method

Participants

Eighteen students at the University of Rochester participated in this experiment and were paid \$7.50. None of them had participated in Experiment 1.

Materials

The materials consisted of 21 pairs of phonologically similar words which differed in lexical frequency according to the Francis and Kučera (1982) counts (e.g., *bell* has a frequency of 23 per million and *bed* a frequency of 139 per million). We were unable to find pairs of words for which one item was very low in frequency and the other very high. However, we made sure that there was a large frequency difference between each item of the pair. The mean frequency of the high- and low-frequency items was 104.0 per million and 14.4 per million, respectively. Each pair was associated with three phonologically unrelated distractors (e.g., *sock, headphones,* and *knife*). The complete set of materials is presented in Appendix C. On a given trial, either the low-frequency or the high-frequency item was presented along with its three distractors. In addition to the 21 pairs and their distractors, 16 sets of four phonologically unrelated items were constructed to serve as filler trials. Two lists were constructed by varying whether the high-frequency or the low-frequency item of each pair was the target. Nine participants were randomly assigned to each list. For each list, three random orders were created; approximately the same number of participants were assigned to each order.

The 169 $[(21 \times 5) + (16 \times 4)]$ pictures were selected from the Snodgrass and Vanderwart (1980) and the Cycowicz, Friedman, Rothstein, and Snodgrass (1997) picture sets as well as from children's picture dictionaries and a commercially available clip-art database; all were black-and-white line drawings.

Before we could conclude that a frequency effect on target fixations was specifically caused by differences in lexical frequency between the two sets of words, we needed to rule out the possibility that a frequency effect might result from differences in object-recognition performances or in name agreement. For instance, people might fixate the pictures associated with low-frequency names more slowly than pictures associated with high-frequency names because the pictured representations of the low-frequency items were harder to recognize or because the words used to refer to the pictures with low-frequency names were less commonly associated with these pictures than were the words and the pictures with high-frequency names. Therefore, we conducted two control experiments. In the first experiment, participants made a speeded semantic categorization on the pictured objects that did not require them to generate item names (i.e., is the pictured object natural or artifact). In the second experiment, they generated a name. The details of these experiments are presented in Appendix D. The results showed no evidence that the pictures of the high-frequency items were recognized more rapidly or named more accurately than the pictures of the low-frequency items. (In fact, the picture-recognition experiment revealed a small bias in favor of the low-frequency item pictures.) Thus, faster fixations to the pictures with higher frequency names could not be attributed to characteristics of the pictures.

The spoken instructions were recorded by a female native speaker of English in a soundproof room, sampling at 22,050 Hz. The average duration of the target word was 625 ms for the high-frequency items and 632 ms for the low-frequency items.

Procedure

As in Experiment 1, participants were seated at a comfortable distance with the computer screen, and their eye movements were monitored as they followed the spoken instructions. The structure of each trial was as follows: First, a 5×5 grid with a centered cross appeared on the screen, and participants were instructed to look at the cross and to click on it. Then four line drawings and four colored geometric shapes appeared on specific cells of the grid. Immediately after the pictures appeared, the spoken instruction started. Because we wanted to analyze the time it took participants to direct their attention to and fixate the target picture, we attempted to minimize the probability for participants to be already fixating the target picture at target onset. The instruction was thus composed of two parts. First, participants were asked to point to one of the distractor pictures using the computer mouse (e.g., ‘‘Point to the sock’’). After a delay of 300 ms to allow participants to move the mouse cursor to the distractor picture, they were instructed to point to the target picture (e.g., ‘‘now the bed’’). Then they were asked to move the target picture above or below one of the geometric shapes (e.g., ‘‘Click on it and put it above the circle’’). Once this was accomplished, the next trial began. By asking participants to point to one of the distractor pictures immediately before pointing to the target picture, we minimized the proportions of trials where subjects were fixating the target picture at the onset of the target word. On 14 of the 16 filler trials, subjects were instructed to click on and move the first picture they pointed to (e.g., ‘‘Point to the balloon. Click on it, and put it below the square’’). This was intended to ensure that people directed their attention to the first picture. The positions of the geometric shapes were fixed from one trial to the other. The position of each picture was randomized for each subject and each trial.

Results and Discussion

The coding procedure was identical to that described for Experiment 1. Fixations were coded from the onset of the target word until participants fixated the target picture and initiated a mouse movement to point to it. One trial was discarded because the participant pointed to the picture without fixating it. Overall, people made 2.14 fixations on average before reaching the target picture, 2.13 fixations when the target was high-frequency, and 2.14 fixations when the target was low-frequency. For 6 trials (1.6% of the data), participants were already fixating the target picture at the onset of the target word; these trials were removed from the subsequent analyses. The delays between the onset of the target and the onset of the last fixation, always to the target, were analyzed after excluding the trials where this delay was less than 200 ms or greater than or equal to 1300 ms. Fixations occurring before 200 ms were assumed to have been programmed before the onset of the target word; trials where participants took more than 1300 ms to fixate the target were treated as outliers. In total, 19 trials (5% of the data) were excluded (12 with low-frequency targets and 7 with high-frequency targets).

On average, the pictures corresponding to high-frequency items were fixated faster than those corresponding to low-frequency items (563 ms vs 625 ms). Planned comparisons revealed that this difference was reliable [$t_1(17) = 3.26$, $p < .005$; $t_2(20) = 1.99$, $p < .05$]. The proportions of fixations to the high-frequency and low-frequency targets over time are shown in Fig.

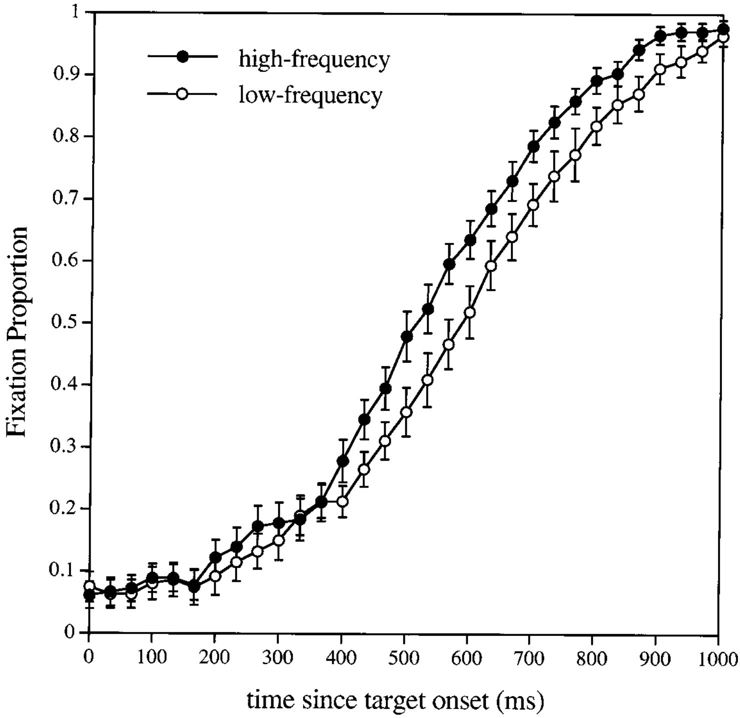

FIG. 8. Experiment 2: Fixation proportions over time for the high-frequency and low-frequency targets. Bars indicate standard errors.

8. At 400 ms after target onset, the proportion of fixations to the high-frequency target surpassed the proportion of fixations to the low-frequency target, indicating that participants fixated the high-frequency target earlier than the low-frequency target. Again, this result reveals the influence of frequency very early in the word-recognition process, given the delay to program and launch an eye movement. The fixation analysis confirms what the analysis on latencies showed: The pictures associated with high-frequency items were fixated faster than the pictures associated with low-frequency items. This result is difficult to reconcile with an account that participants in our experiments adopted a special verification strategy based on frequency. We discuss the implications of this result for the eye-tracking paradigm in more detail under General Discussion. We also discuss additional empirical evidence that demonstrates that word recognition in the eye-tracking paradigm does not bypass normal lexical processing.

Simulations

The same procedures were used in the present simulations as were used for Experiment 1. However, for each pair of high- and low-frequency targets, each target word was presented as a separate input to TRACE. Three distractors with similar frequencies to those for the distractors used in the experiment were chosen from the TRACE lexicon. The same three distractors were used for both members of each pair. Figure 9 presents the overall activation patterns. In the top panel, average high- and low-frequency target activations are shown when frequency was not used. The middle panel shows the corresponding averages observed when frequency was incorporated in resting activations (frequency scaling constant $s = .06$), and the bottom panel shows the results when frequency was incorporated in phoneme-to-word connection strengths (frequency scaling constant $s = .13$). Note that there was a slight advantage for high-frequency items when frequency was not included. This was due to differences in neighborhoods among the items. However, the additional difference due to the effect of frequency was clearly substantial.

In contrast to Experiment 1, the task used in Experiment 2 was similar to that used by Allopenna et al. (1998). While Allopenna et al. used explicit fixation instructions to ensure that subjects fixated a single location at target onset, we devised task constraints that made it likely that subjects would not be fixating the target picture at the onset of the target word (i.e., by asking them to point to a distractor picture before the critical instruction). Therefore, for Experiment 2, an additional method for computing response probabilities was included: the scaled Luce choice rule used by Allopenna et al. As discussed above, the simple form of the Luce choice rule predicts that for equivalent response strengths (even when all are 0), the minimum response probability is $1/n$, where n is the number of possible responses. In order to approximate a context in which task constraints dictate nonuniform probabilities for all responses (e.g., when subjects are instructed to fixate a particular item or to make a visually guided movement to an object other than the target prior to the critical instruction), Allopenna et al. scaled activations as a function of activation in the current time slice relative to the maximum activation observed during processing, using the scaling factor shown in Eq. (7), where $\max(a_t)$ is the maximum activation in time slice t, and $\max(a)$ is the maximum activation observed during all time steps:

$$\Delta = \frac{\max(a_t)}{\max(a)}. \tag{7}$$

We incorporated this into our choice rule simply by multiplying response strengths by the scaling factor prior to computing response probabilities. We refer to response probabilities computed with this method as $\mathrm{AMT}_{\mathrm{NO\ FRQ}}$ (Allopenna et al. method, no frequency), $\mathrm{AMT}_{\mathrm{REST}}$ (activations with frequency instantiated in resting levels), and $\mathrm{AMT}_{\mathrm{WT}}$ (activations with frequency instantiated in connection weights).

In order to compute scaled response probabilities with frequency instanti-

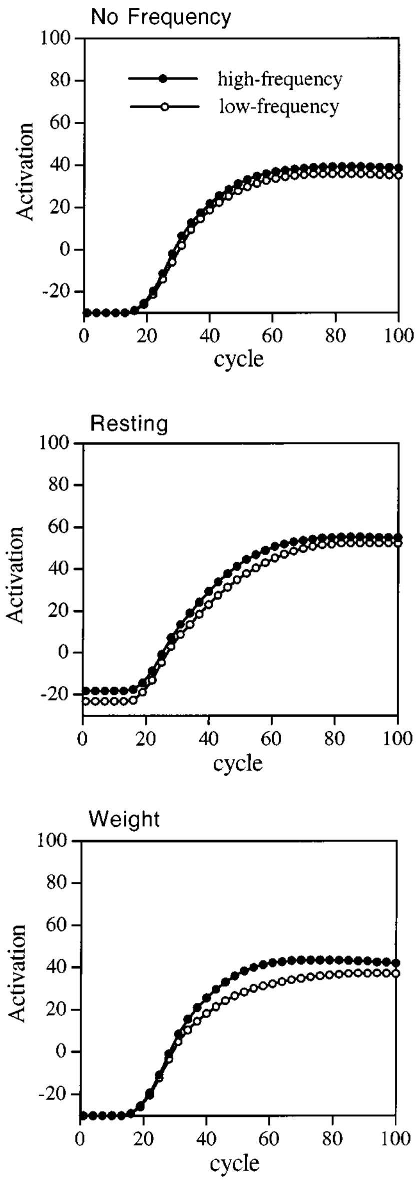

FIG. 9. Experiment 2 simulations: TRACE activations over time for the high- and low-frequency targets when frequency is turned off (top), when frequency is coded in resting activations (middle), and when frequency is coded in connection weights (bottom).

ated at a ‘‘postactivation’’ stage (AMT_{POST}), we needed a different scaling factor than the one described in Eq. (7). This is because the scaling factor in Eq. (7) is based on activations; in the decision-rule method, raw activations do not incorporate frequency. In order to incorporate frequency in the scaling without directly applying it on activations, we devised the method presented in Eq. (8):

$$\Delta = \frac{\max(a_t + f^4)}{\max(a + f^4)}, \tag{8}$$

where a_t is the maximum activation (without frequency) at time *t; a* the maximum activation observed during all time steps; and *f* the item’s log-transformed frequency [see Eq. (6)] raised to the fourth power (to amplify the frequency effect).

Figure 10 presents the human data for the high- and low-frequency targets, as well as response probabilities calculated using the four different methods used for Experiment 1 ($RDL_{NO\ FRQ}$, RDL_{REST}, RDL_{WT}, and RDL_{POST}). In Figure 11, we present the data with the AMT methods computed using activations without frequency ($AMT_{NO\ FRQ}$), with frequency in resting levels (AMT_{REST}), in connection weights (AMT_{WT}), or in a scaled decision rule (AMT_{POST}). Figure 12 superimposes the differences between high-frequency and low-frequency targets from the data and the four RDL and AMT variants.

Tables 3 and 4 present RMS and r^2 values for the high- and low-frequency targets and for the difference between the two for two windows of interest: a window including all data points from 0 to 1000 ms and a window included data points from 200 to 1000 ms, since the frequency difference manifests itself from about 200 ms onward.

The fit measures confirm the impression given by Fig. 12. As for Experiment 1, the RDL_{WT} method is superior to RDL_{REST} and RDL_{POST}, and the AMT methods, by explicitly incorporating task constraints, provide the best fit to the data. The three AMT methods that incorporate frequency (AMT_{REST}, AMT_{WT}, and AMT_{POST}) provide roughly comparable fits as measured by RMS and r^2 values for the high-frequency/low-frequency difference for both window analyses. AMT_{POST} provides high r^2 values, but its RMS is 30% higher than those for the other two AMT variants. This is because the scaling factor in Eq. (8) only partially succeeds in reducing initial activation values. Most of the improvement in fit compared to RDL_{POST} is due to sustaining the high-frequency/low-frequency difference. Thus, a substantial contribution toward fitting this data comes from choice rules that explicitly incorporate the task constraints faced by the subject.

These simulations replicate the main result of the simulations of Experiment 1 using the same set of parameters. Response probabilities computed from TRACE activations account for substantial amounts of variance in our

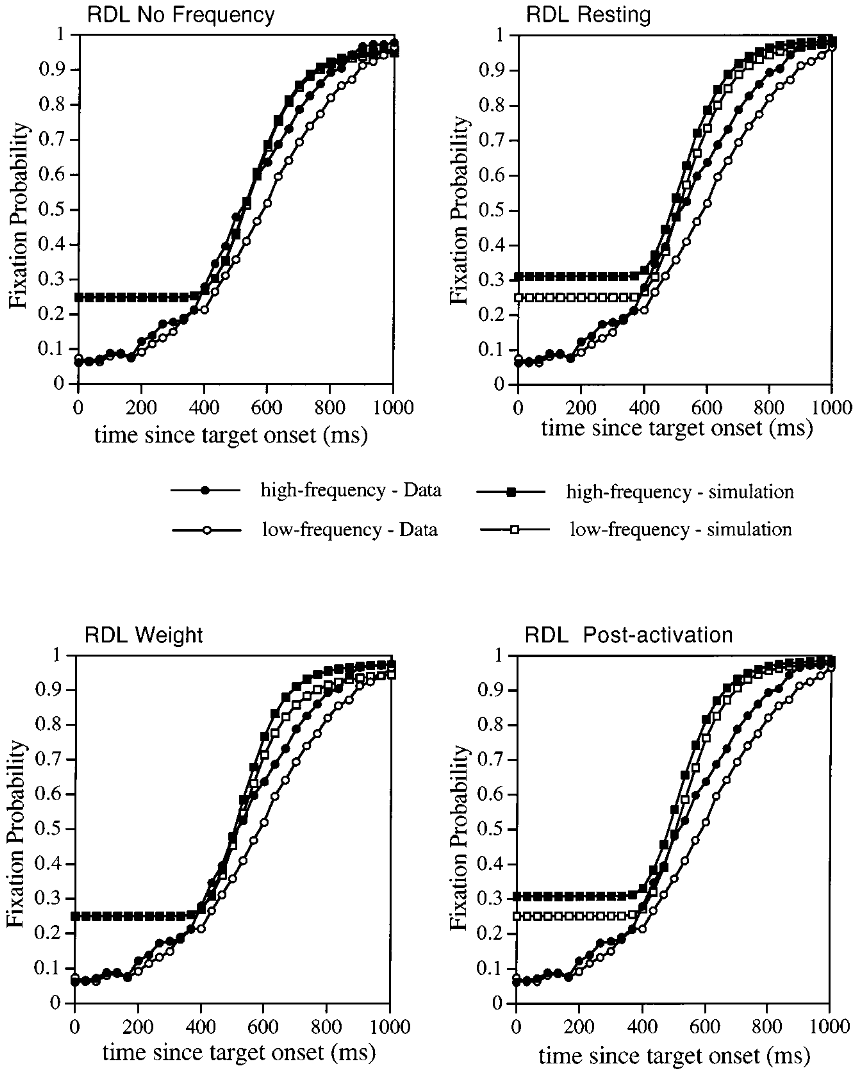

FIG. 10. Experiment 2 simulations: Fixation probabilities over time for the target, the high- and low-frequency competitors, and the distractor, for the data and each of the four RDL frequency implementations (see text).

time-course data, and the different methods of implementing frequency do not differ greatly. The importance of this result will be discussed under General Discussion. The simulations of the data from Experiment 2 emphasize the importance of representing task constraints in simulations. The AMT methods, which approximate the task constraints of Experiment 2, provide better fits than the RDL methods.

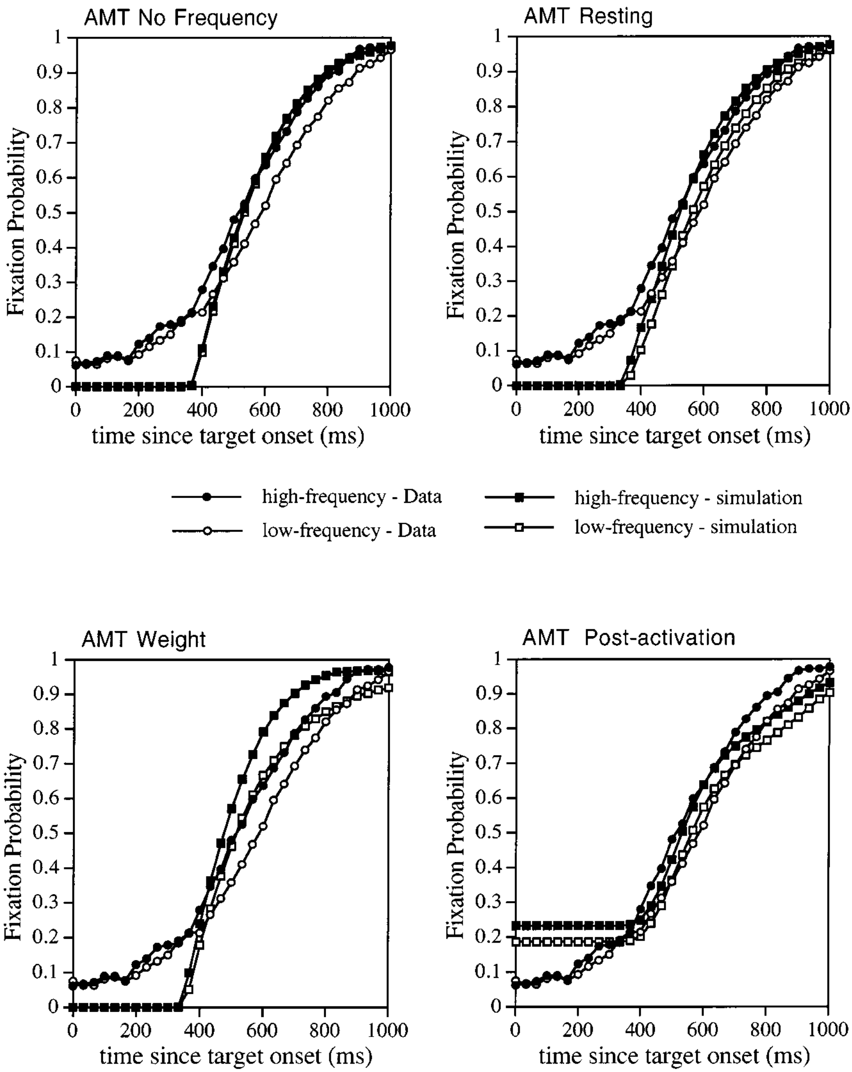

FIG. 11. Experiment 2 simulations: Fixation probabilities over time for the target, the high- and low-frequency competitors, and the distractor for the data and each of the four AMT frequency implementations (see text).

GENERAL DISCUSSION

Two experiments used eye movements to investigate the time course of frequency effects during spoken-word recognition in continuous speech using a task in which subjects followed spoken instructions to click on and move pictures in a display. In Experiment 1, a target picture was presented

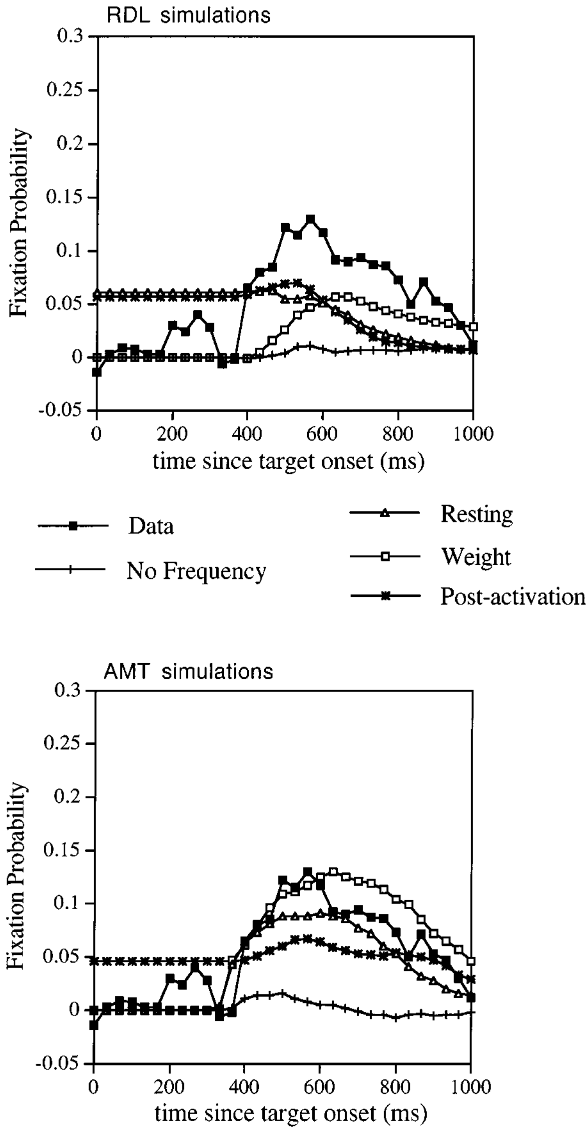

FIG. 12. Experiment 2 simulations: Fixation-probability differences between the high- and the low-frequency competitors over time for the data and each of the four RDL frequency implementations (top) and the data and each of the four AMT frequency implementations (bottom).

TABLE 3
RMS Measures of Model Fits to Experiment 2 for the Target, the High-Frequency (HF) and the Low-Frequency (LF) Cohort Competitors, and the Distractor on the 0- to 1000-ms (All) and the 200- to 1000-ms (Section) Windows

Mean	Item	$RDL_{NO\ FRQ}$	RDL_{REST}	RDL_{WT}	RDL_{POST}	$AMT_{NO\ FRQ}$	AMT_{REST}	AMT_{WT}	AMT_{POST}
.492	HF Target (All)	.092	.141	.105	.145	.092	.084	.104	.086
.592	HF Target (Section)	.062	.106	.080	.093	.096	.085	.109	.058
.440	LF Target (All)	.121	.140	.128	.149	.102	.082	.095	.066
.527	LF Target (Section)	.078	.130	.113	.111	.108	.083	.099	.048
.054	HF–LF (All)	.063	.051	.042	.049	.064	.022	.024	.036
.065	HF–LF (Section)	.070	.048	.047	.045	.072	.025	.027	.034

Note. Means for the data are shown in the leftmost column to guide interpretation of the RMS values.

TABLE 4
r^2 Measures of Model Fits to Experiment 2 for the Target, the High-Frequency (HF) and the Low-Frequency (LF) Cohort Competitors, and the Distractor on the 0- to 1000-ms (All) and the 200- to 1000-ms (Section) Windows

Item	$RDL_{NO\ FRQ}$	RDL_{REST}	RDL_{WT}	RDL_{POST}	$AMT_{NO\ FRQ}$	AMT_{REST}	AMT_{WT}	AMT_{POST}
HF Target (All)	.967	.968	.966	.967	.987	.990	.966	.974
HF Target (Section)	.960	.972	.969	.964	.989	.990	.950	.985
LF Target (All)	.971	.965	.966	.960	.970	.985	.953	.982
LF Target (Section)	.974	.958	.960	.965	.960	.982	.933	.985
HF–LF (All)	.424	.032	.570	.002	.212	.824	.790	.614
HF–LF (Section)	.179	.004	.436	.067	.229	.760	.695	.689

along with two pictures associated with cohort competitors and a distractor picture. The cohort competitors differed in frequency. As the target word unfolded over time, the competitors were fixated more than the distractor, replicating ‘‘cohort’’ effects obtained in previous work (e.g., Allopenna et al., 1998). Crucially, the high-frequency competitor was more likely to be fixated than the low-frequency competitor. Experiment 2 extended this result by demonstrating an effect of frequency independent of any possible competition from the displayed competitors. We presented a referent along with three phonologically unrelated pictures; the referent was either low- or high-frequency. The latency with which participants fixated the referent picture was shorter for high-frequency targets than for low-frequency targets. Simulations using activations generated from the TRACE model of spoken-word recognition and the linking hypothesis introduced by Allopenna et al. (1998) provided close fits to the data. These simulations provide strong support for the claim that frequency influences even the earliest moments of lexical access. Frequency implementations within the TRACE model provided excellent fits to the data regardless of whether frequency was instantiated in resting-activation levels or connection weights or as a bias applied to activations at each time slice. Implementations of frequency in connection weights provided better qualitative and quantitative fits than the other methods.

The results have both theoretical and methodological implications for our understanding of spoken-word recognition. On the theoretical side, our data demonstrate that frequency has immediate effects on spoken-word processing and these effects persist for a substantial amount of time beyond the offset of a stimulus. Thus, the results strongly support models in which frequency is intrinsic to the word-recognition process and affects lexical activation. We found no evidence that frequency effects are delayed until activation has passed some threshold. Rather, frequency effects are present at the earliest measurable points of the recognition process, certainly long before there is sufficient information for a decision stage to make reliable decisions. This finding in itself does not refute the argument that characteristics of some tasks amplify or reduce the influence of lexical frequency on performances (Balota & Chumbley, 1984, 1985; Sommers et al., 1997) or that such task-based influences operate via a late response bias. This research shows that the influence of *intrinsic* lexical frequency (i.e., the frequency with which sound patterns occurs in the language) operates very early during spoken-word recognition. This suggests that models of spoken-word recognition (such as the NAM) will account for increased variance when they incorporate metrics that calculate frequency effects using dynamically changing competitor sets. Our simulations show that the time course of frequency effects closely follows predictions from continuous mapping models, such as TRACE. Moreover, a variety of frequency implementations and choice rules can account for these time-course effects. More specifically, tests of three different frequency implementations (frequency as a bias affecting resting-activation level, frequency as a variable that influences connection strength, and frequency as a postactivation bias) showed that all three give roughly equivalent accounts of the data, suggesting that the crucial question is not exactly how frequency operates, but how immediate frequency effects are. The present work, by translating qualitative models into formal ones, reveals that the two approaches to frequency most often sharply contrasted in the literature, the resting-activation level and postactivation bias accounts, make strikingly similar predictions. This is not as surprising as it might seem; in Bayesian terms, both of these models can be viewed as implementing frequency as a bias based on priors.

Although the three implementations of frequency used here yielded very good fit to the data, the connection-strength approach captured the overall morphology of the frequency effects somewhat better than the other two models. Overall, the connection-strength approach provided the best fits for the data from Experiment 1, especially the differences between the high- and low-frequency competitors. This was because the resting-activation and postactivation models predicted baseline frequency differences that were not present in the data. The connection-weight simulations also provided the best fits for Experiment 2 for the RDL simulations, but not for the AMT simulations, for reasons discussed above. The connection-strength account is also easiest to reconcile with models in which frequency (and contingent frequency) reflects experience accumulated over time as a result of learning. Finally, the connection-weight approach is also most compatible with models in which frequency might have effects at multiple levels of representations, e.g., joint effects of lexical frequency and phonotactics. However, it is possible that the other models could be improved with some parameter adjustments.

Given our results, it is important to reconsider the evidence that has been used to argue against a resting-activation account of frequency. We have already discussed some of the limitations with the Connine et al. (1993) arguments against a resting-activation implementation of frequency. The second empirical argument comes from studies examining phonological priming. Goldinger et al. (1989) reported trends for low-frequency primes to generate stronger inhibitory priming than high-frequency primes. They used the NAM to explain this result. High-frequency items reach recognition threshold more quickly than low-frequency items. Once an item is recognized, the system is hypothesized to be reset and word activations to begin to decay back to resting levels. As a result, activation of the acoustic-phonetic pattern corresponding to a high-frequency prime will begin to return to a resting level sooner than activation of a low-frequency prime. The result, given equivalent delays until the presentation of a subsequent target, is greater competition from residual activation due to a low-frequency prime. According to Goldinger et al., a model in which frequency operates directly on activation levels (either on resting activations or connection weights) would predict more inhibition from high-frequency primes than from low-frequency primes because high-frequency primes should produce stronger competing activation. However, Luce, Goldinger, Auer, and Vitevitch (2000) reported simulations with a connectionist model (''PARSYN'') in which frequency is instantiated in connection strengths (i.e., the weights of lateral connections between phonemes are proportional to forward and backward transitional probabilities), which correctly predicted the trend for stronger priming by low-frequency items. As Luce et al. note, PARSYN is similar to a class of models, including TRACE, which presumably would make similar predictions. Thus, the reported priming difference between low- and high-frequency primes does not depend on frequency operating as a postactivation bias.

Our results also have important methodological implications for eye-movement studies of spoken-word recognition. The eye-tracking methodology is extremely sensitive to the time course of lexical processing; it can be used with continuous speech without interrupting the speech stream and does not require that participants make an overt decision. Moreover, given an explicit model, hypothetical activation patterns (or more generally degrees of evidence) can be mapped onto predicted fixation patterns using an explicit linking hypothesis. Our results and simulations suggest that the approach adopted here can be used to develop and evaluate models of the temporal dynamics of lexical access. It is particularly encouraging that the approach adopted by Allopenna et al. (1998) provided excellent fits to the data reported here, using the same parameters and principled mapping of processing cycles onto recognition time, despite differences in procedure. Moreover, the only changes to the simulations between Experiments 1 and 2 were changes that reflected the slightly different task constraints.

An inherent constraint on the eye-tracking methodology is that the spoken instruction can only make reference to a restricted set of objects that are visually available. This raises the concern that subjects might adopt a special verification strategy that bypasses normal lexical processing. This seems highly unlikely given the results presented here and the procedure we adopted to reduce preview time. If the lexical candidates that entered the recognition process were restricted to the visually present alternatives, we would not expect to see effects of frequency. This is especially true for Experiment 2, where we found clear frequency effects even when the display did not contain any competitors with names that were similar to the referent.

However, it is possible to argue that the only unequivocal evidence for effects from the general lexicon would be evidence that fixations to a referent are influenced by lexical competitors that are neither named nor pictured. Dahan, Magnuson, Tanenhaus, and Hogan (in press) provided just such a demonstration in a recent study that examined the time course of lexical competition when mismatching coarticulatory information was created by cross-splicing two different speech fragments. Participants were slower to fixate the target picture (e.g., the picture of a net) when the onset of the target word originated from a competitor word (e.g., *ne*[ck] + [ne]*t*) than from a nonword (e.g., *ne*[p] + [ne]*t*). Crucially, this effect was found even when the competitor word (e.g., *neck*) was neither named nor pictured during the experiment. This demonstrates a clear effect of lexical competition that cannot be accounted for by any form of verification strategy. Moreover, we were able to model the time course of the mismatch effects using the same parameters as we used in the current simulations.

In summary, the present experiments and simulations make the following contributions. First, the time course of the experimental results rules out models in which frequency effects in spoken-word recognition are primarily due to decision biases that apply after lexical activation is complete. Second, the simulations demonstrate that (a) models incorporating frequency provide better quantitative fits for lexical activations than models without frequency; (b) models incorporating frequency as a change in resting activation and models incorporating frequency as a continuously operating response bias make similar predictions, despite the fact that they are often opposed in the literature; and (c) models treating frequency as changes in connection weights capture some aspects of the data that the other models do not.

The temporal sensitivity of the eye-tracking paradigm, when coupled with the use of an explicit linking hypothesis, makes it a powerful tool for investigating the time course of spoken-word recognition and for evaluating predictions made by alternative models. The current results suggest that the paradigm should be applicable to a wide range of questions about the time course of lexical access.

APPENDIX A
Experiment 1

Target set A	HF cohort	LF cohort	Target set B	Trace T	Trace HF	Trace LF
band aid/bændeɪd/(0)	bank/bæŋk/(110)	banjo/bændʒəʊ/(1)	stove	/bɒbdɒd/	/bɒbk/	/bɒbʃu/
bench/bentʃ/(42)	bed/bed/(139)	bell/bel/(23)	mushroom	/bibʃ/	/bid/	/bil/
cane/keɪn/(13)	cake/keɪk/(16)	cage/keɪdʒ/(11)	ladder	/kug/	/kuk/	/kuʃ/
cannon/kænən/(4)	camera/kæmɹə/(46)	camel/kæməl/(2)	bolt	/kɒbʌg/	/kɒbʌɹʌ/	/kɒbʌl/
carpet/kaɹpɪt/(17)	card/kaɹd/(61)	cart/kaɹt/(9)	helmet	/kɒɹpʌt/	/kɒɹd/	/kɒɹt/
cellist/tʃelɪst/(2)	chair/tʃeəɹ/(89)	chess/tʃes/(3)	rabbit	/ʃʌlist/	/ʃʌiɹ/	/ʃʌs/
chimney/tʃɪmnɪ/(7)	children/tʃɪldrən/(346)	chicken/tʃɪkən/(49)	racket	/ʃibpi/	/ʃildɹʌg/	/ʃikʌg/
clover/kləʊvəɹ/(16)	cloud/klaʊd/(64)	clown/klaʊn/(6)	grasshopper	/klugʌɹ/	/klud/	/klub/
coconut/kəʊkənʌt/(10)	coat/kəʊt/(52)	cone/kəʊn/(15)	pumpkin	/kukubʌt/	/kut/	/kun/
hanger/hæŋəɹ/(1)	hand/hænd/(717)	hammer/hæməɹ/(6)	butterfly	/sɒpgʌɹ/	/sɒpd/	/sɒbʌɹ/
mountain/maʊntɪn/(98)	mouth/maʊθ/(113)	mouse/maʊs/(20)	helicopter	/buptʌb/	/buʃ/	/bus/
pizza/pitsə/(3)	peanut/pinʌt/(11)	peacock/pikɒk/(6)	brain	/pʌsɒ/	/pʌbʌt/	/pʌkuk/
pepper/pepəɹ/(13)	pencil/pensəl/(38)	penguin/peŋgwɪn/(0)	axe	/pupʌɹ/	/pubsil/	/pubgib/
pillow/pɪləʊ/(11)	picture/pɪktʃəɹ/(227)	pickle/pɪkəl/(2)	canoe	/pilu/	/piktɹ/	/pikl/
snowman/snəʊmæn/(0)	snake/sneɪk/(70)	snail/sneɪl/(3)	watch	/spubʌb/	/spɒk/	/spɒl/
trumpet/tɹʌmpɪt/(6)	truck/tɹʌk/(80)	trunk/tɹʌŋk/(13)	pen	/tɹʌbʌt/	/tɹʌk/	/tɹʌ^bk/
whistle/wɪsəl/(3)	window/wɪndəʊ/(172)	windmill/wɪndmɪl/(1)	squirrel	/ʃusʌl/	/ʃubdu/	/ʃubdbil/

Note: Lexical frequency as reported in Francis and Kučera (1982) is indicated in parentheses (a frequency of 0 indicates that the word was not found in the Francis and Kučera database). In addition to the IPA transcription of the targets, high-frequency, and low-frequency cohort items, we indicate the TRACE transcription of these words used in the simulations (given the limited phoneme inventory available in TRACE).

APPENDIX B
TRACE Parameters

Maximum activation	1.0
Minimum activation	−0.3
Feature-level decay	0.01
Phoneme-level decay	0.03
Word-level decay	0.05
Feature-to-phoneme excitation	.02
Phoneme-to-word excitation	.05
Word-to-phoneme excitation	.03
Phoneme-to-feature excitation	.00
Feature-level inhibition	0.04
Phoneme-level inhibition	0.04
Word-level inhibition	0.03
Feature default resting activation	−0.10
Phoneme default resting activation	−0.10
* Word default resting activation	−0.30 [−0.01]
* Word printing threshold	−1 [.05]
Feature continua weights: all set to 1.0	
* *F* scale (the frequency scaling constant *s*):	
0 (No-Frequency simulations) [0.0]	
0.06 (Resting-activation simulations)	
0.13 (Connection-weight simulations)	

Note. An asterisk indicates parameters set to something other than the original TRACE defaults. The number in square brackets indicates the original value.

APPENDIX C
Experiment 2

Low frequency	High frequency	Distractor 1	Distractor 2	Distractor 3	Trace HF	Trace LF
bell/bel/(23)	bed/bed/(139)	sock	headphones	knife	/bil/	/bid/
cap/kæp/(22)	cat/kæt/(42)	leaf	fly	wheel	/kɒp/	/kɒt/
cart/kaɹt/(9)	card/kaɹd/(61)	tiger	whistle	globe	/kɒɹt/	/kɒɹd/
chain/tʃeɪn/(60)	chair/tʃeəɹ/(89)	buffalo	trumpet	carrot	/ʃib/	/ʃiɹ/
chicken/tʃɪkən/(49)	children/tʃɪldɹən/(346)	lock	pear	ashtray	/ʃikʌg/	/ʃildɹʌg/
clown/klaʊn/(6)	cloud/klaʊd/(64)	shark	grill	telephone	/klub/	/klud/
comb/kəʊm/(6)	coat/kəʊt/(52)	onion	bird	necklace	/kub/	/kut/
couch/kaʊtʃ/(2)	cow/kaʊ/(46)	pumpkin	jar	strawberry	/kʌʃ/	/kʌʌ/
flask/flɑsk/(5)	flag/flæg/(18)	deer	avocado	iron	/dlɒsk/	/dlɒg/
harp/haɹp/(1)	heart/haɹt/(199)	umbrella	giraffe	duck	/lɒɹp/	/lɒɹt/
horn/hɔɹn/(33)	horse/hɔɹs/(233)	sun	plug	kettle	/liɹb/	/liɹs/
lamb/læm/(14)	lamp/læmp/(24)	orange	gun	ear	/lɒb/	/lɒbp/
mouse/maʊs/(20)	mouth/maʊθ/(113)	anchor	basket	desk	/bus/	/buʃ/
penguin/peŋgwɪn/(0)	pencil/pensəl/(38)	grapes	ladybug	binoculars	/pubgib/	/pubsil/
pitcher/pɪtʃəɹ/(29)	picture/pɪktʃəɹ/(277)	tie	kite	banana	/pitʃɹ/	/piktɹ/
plane/pleɪn/(2)	plate/pleɪt/(44)	scissors	well	camel	/plɒb/	/plɒt/
snail/sneɪl/(3)	snake/sneɪk/(70)	zipper	rocking chair	glasses	/spɒl/	/spɒk/
trunk/tɹʌŋk/(13)	truck/tɹʌk/(80)	butterfly	watch	ring	/tɹʌbk/	/tɹʌk/
skate/skeɪt/(1)	scale/skeɪl(62)	turtle	moon	peanut	/skit/	/skil/
toad/təʊd/(4)	toe/təʊ/(14)	book	pineapple	feather	/tʌd/	/tʌʌ/
windmill/wɪndmɪl/(1)	window/wɪndəʊ/(172)	pipe	turkey	pen	/ʃubdbil/	/ʃubdu/

Note. Lexical frequencies for the high-frequency and low-frequency items are in parentheses; IPA transcriptions of these items, as well as the TRACE transcriptions of these words, are indicated.

APPENDIX D

Control Experiments for Experiment 2

Ten participants (who did not participate in Experiment 2) took part in the control experiments. First, participants were presented with the pictures of the 21 pairs and their distractors on a computer screen and had to decide, for each picture successively, whether the object represented exists in nature or is artificial (i.e., ‘‘man-made’’). Participants were instructed to decide as quickly and accurately as possible and press one of two keys of the computer keyboard to indicate their answer. A few practice trials using distractor pictures initiated the session. Accuracy and reaction times were collected. This task required the participants to recognize the object in order to make their decision without requiring access to its name. If pictures for high-frequency and low-frequency items could be recognized equally well, we expected to see comparable decision times for both types. However, some pictures were difficult to classify [e.g. *clown* and *heart* (represented by the card-suit symbol)] and yielded incorrect responses. After excluding all incorrect responses (3.6% of the data; 11 on high-frequency items and 12 on low-frequency items), we analyzed the reaction times. High-frequency pictures were responded to more slowly than low-frequency pictures [749 ms vs 674 ms, $t(9) = 2.79$, $p < .05$]. This effect was also found when the corrected responses from *heart* and *clown* were excluded [$t(9) = 3.14$, $p < .05$]. Thus, the pictures associated with the low-frequency items appeared to be recognized more easily than the pictures associated with the high-frequency items. If this difference were to play a role in the eye-tracking study, it would go against the predicted advantage in the latency for fixating high-frequency target pictures compared to low-frequency target pictures.

The second control experiment consisted in subsequently presenting the same participants with the same pictures and asking them to name each picture by typing the name on the computer keyboard. Name agreement for the high-frequency and low-frequency pictures did not differ significantly (84.3% vs 82.9%, $t < 1$). Name agreement for *heart* and *clown* was very high (90 and 100%, respectively), indicating that the incorrect responses observed in the picture-recognition experiment resulted from the nature of the classification, not from the pictures themselves.

These control experiments indicated that the pictures associated with high-frequency items were not easier to recognize or more accurately named than the pictures associated with the low-frequency items. Thus, faster fixations to the pictures with high-frequency names could not be attributed to characteristics of the pictures.

REFERENCES

Allen, P. A., McNeal, M., & Kvak, D. (1992). Perhaps the lexicon is coded as a function of word frequency. *Journal of Memory and Language,* **31,** 826–844.

Allen, P. A., Smith, A. F., Lien, M.-C., Weber, T. A., & Madden, D. J. (1997). Word frequency effects at brief exposure durations: Comment on Paap and Johansen (1994). *Journal of Experimental Psychology: Human Perception and Performance,* **23,** 1792–1797.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language,* **38,** 419–439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition,* **73,** 247–264.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance,* **10,** 340–357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language,* **24,** 89–106.

Balota, D. A., & Chumbley, J. I. (1990). Where are the effects of frequency in visual word recognition tasks? Right where we said they were! Comment on Monsell, Doyle, and Haggard (1989). *Journal of Experimental Psychology: General,* **119,** 231–237.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **19,** 81–94.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology,* **6,** 84–107.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology,* **65,** 171–237.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (in press). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes.*

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes,* **5,** 313–349.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research,* **24,** 409–436.

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 31–45). New York: Springer-Verlag.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance,* **10,** 526–540.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Frauenfelder, U. H., & Peeters, G. (1998). Simulating the time course of spoken word recognition: An analysis of lexical competition in TRACE. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 101–146). Mahwah, NJ: Erlbaum.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes,* **12,** 613–656.

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,** 501–518.

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

Grosjean, F., & Frauenfelder, U. (1996). *A guide to spoken word recognition paradigm.* Hove: Psychology Press.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology,* **41,** 401–410.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science,* **11,** 32–38.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–301). San Diego: Academic Press.

Luce, P. A. (1986). *Neighborhood of words in the mental lexicon.* Research on speech perception, Technical report No. 6. Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics,* **62,** 615–625.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing,* **19,** 1–36.

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

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review,* **89,** 483–506.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills.* New York: Springer-Verlag.

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

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148–172). Hove: Erlbaum.

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

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance,* **13,** 14–24.

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

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. I. An account of basic findings. *Psychological Review,* **88,** 375–407.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance,* **17,** 433–443.

McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in word naming. *Journal of Memory and Language,* **29,** 43–65.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Bresner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Erlbaum.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118,** 43–71.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences,* **23,** 299–370.

Paap, K. R., & Johansen, L. S. (1994). The case of the vanishing effect: a retest of the verification model. *Journal of Experimental Psychology: Human Perception and Performance.* **20,** 1129–1157.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition. *Psychological Review,* **89,** 573–594.

Plaut, D. C., McClelland, J., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review,* **103,** 56–115.

Pollack, I., Rubenstein, H., & Decker, L. (1959). Intelligibility of known and unknown message sets. *The Journal of Acoustic Society of America*, **31,** 273–279.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America,* **57,** 1030–1033.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review,* **96,** 523–568.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory,* **6,** 174–215.

Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). Some consideration in evaluating spoken word recognition by normal-hearing, noise-masked normal hearing, and cochlear implant listeners. I. The effects of response format. *Ear and Hearing,* **18,** 89–99.

Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language and Cognitive Processes*, **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,** 1632–1634.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition,* **73,** 89–134.

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics,* **36,** 217–222.

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

(Accepted March 29, 2000)