Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 5, 1108–1117

0278-7393/10/$12.00 DOI: 10.1037/a0019901

Tracking the Time Course of Orthographic Information in Spoken-Word Recognition

Anne Pier Salverda and Michael K. Tanenhaus
University of Rochester

Two visual-world experiments evaluated the time course and use of orthographic information in spoken-word recognition using printed words as referents. Participants saw 4 words on a computer screen and listened to spoken sentences instructing them to click on one of the words (e.g., *Click on the word bead*). The printed words appeared 200 ms before the onset of the spoken target word. In Experiment 1, the display included the target word and a competitor with either a lower degree (e.g., *bear*) or a higher degree (e.g., *bean*) of phonological overlap with the target. Both competitors had the same degree of orthographic overlap with the target. There were more fixations to the competitors than to unrelated distractors. Crucially, the likelihood of fixating a competitor did not vary as a function of the amount of phonological overlap between target and competitor. In Experiment 2, the display included the target word and a competitor with either a lower degree (e.g., *bare*) or a higher degree (e.g., *bear*) of orthographic overlap with the target. Competitors were homophonous and thus had the same degree of phonological overlap with the target. There were more fixations to higher overlap competitors than to lower overlap competitors, beginning during the temporal interval where initial fixations driven by the vowel are expected to occur. The authors conclude that orthographic information is rapidly activated as a spoken word unfolds and is immediately used in mapping spoken words onto potential printed referents.

Keywords: spoken-word recognition, orthography, phonology, visual-world paradigm, eye movements

One of the more surprising findings in research on spoken-word recognition is that orthographic representations become available upon hearing a spoken word (see Frost & Ziegler, 2007, for a review). An early study by Seidenberg and Tanenhaus (1979) found that rhyme judgments for pairs of spoken words were delayed for orthographically dissimilar words (e.g., *tie*, *rye*) compared to orthographically similar words (e.g., *tie*, *lie*). Orthographic information is not relevant for making rhyme judgments, and therefore, a priori, one would not expect to find evidence for the activation of orthographic information in this task. Initially, there were concerns that this result might be due to strategic processing of the stimuli (but cf. Donnenwerth-Nolan, Tanenhaus, & Seidenberg, 1981; Tanenhaus, Flanigan, & Seidenberg, 1980). However, later studies have employed paradigms using less explicit response measures. For instance, Ziegler and Ferrand (1998) demonstrated that in an auditory lexical decision task, participants make slower responses to words that are orthographically inconsistent (i.e., words whose rhyme can be spelled in multiple ways, e.g., *beak*) than to words that are orthographically consistent (i.e., words whose rhyme can be spelled in only one way, e.g., *luck*). Orthographic effects have been found in a variety of other tasks, such as primed lexical decision (Chereau, Gaskell, & Dumay, 2007; Perre, Midgley, & Ziegler, 2009), pseudohomophone priming (Taft, Castles, Davis, Lazendic, & Nguyen-Hoan, 2008), semantic categorization (Pattamadilok, Perre, Dufau, & Ziegler, 2009), and gender categorization (Peereman, Dufour, & Burt, 2009). Taken together, these results show that orthographic effects in spoken-word recognition are not strategic.

Two recent event-related potential (ERP) studies were the first to provide information concerning the time course of the activation of orthographic information in spoken-word recognition (Pattamadilok et al., 2009; Perre & Ziegler, 2008). In Perre and Ziegler's (2008) study, French participants heard words that were either orthographically consistent or orthographically inconsistent. Perre and Ziegler further manipulated the location of the inconsistency: Orthographically inconsistent words had an early inconsistency (i.e., their body could be spelled in multiple ways) or a late inconsistency (i.e., their rhyme could be spelled in multiple ways). The lexical decision results replicated and extended the findings of Ziegler and Ferrand (1998), with both types of orthographically inconsistent words being responded to more slowly than orthographically consistent words. Perre and Ziegler also observed differences in the magnitude of an ERP component, dubbed the N320, which was more negative in the early inconsistency condition than in the consistent condition. Taking the consistent condition as a baseline, differences in ERP signal were observed from about 200 ms after the onset of the vowel in the early inconsistency condition. This difference in negativity of the ERP signal thus appears to be closely time locked to the processing of the rhyme of the target word. The ERP results for the late inconsistency condi-

Anne Pier Salverda and Michael K. Tanenhaus, Department of Brain and Cognitive Sciences, University of Rochester.

This research was partially supported by National Institute on Deafness and Other Communication Disorders Grant DC005071. We thank Mary Hare and Mark Seidenberg for helpful comments on this research. We are grateful to Neil Bardhan for recording the stimuli and to Austin Frank for statistical consultation.

Correspondence concerning this article should be addressed to Anne Pier Salverda, Department of Brain and Cognitive Sciences, University of Rochester, Meliora Hall, Box 270268, Rochester, NY 14627-0268. E-mail: asalverda@bcs.rochester.edu

tion also showed a more pronounced negativity, at around 200 ms after the offset of the target word, although this effect was weaker than the ERP effect in the early inconsistency condition. These results, and similar results by Pattamadilok et al. (2009) using a semantic categorization task, provided the first evidence concerning the time course of the activation of orthographic information in spoken-word recognition, and they suggest that such information is activated with very little delay, as a word is heard and processed.

Taken together, then, there is abundant evidence from studies using a variety of experimental paradigms that orthographic information can influence a listener's performance in an experimental task where such information is not relevant. These results demonstrate convincingly that some form of orthographic representation is automatically activated as a consequence of hearing a spoken word. However, it is unclear what role, if any, is served by these orthographic representations. It is usually assumed that orthographic representations play no functional role in language comprehension. According to this view, spoken words (and sounds therein) readily activate orthographic representations simply because those representations have become strongly associated with a spoken word as a consequence of abundant practice in reading and spelling. A related view is that the process of learning to read actually changes the nature of phonological representations. Shifts in phonological representations emerge as a natural by-product of attractor and feedback mechanisms in parallel distributed processing (PDP) models, such as Harm and Seidenberg (2004).

In this article, we evaluate a hypothesis about a functional role for orthographic representations in spoken-word recognition that has not, to the best of our knowledge, been previously considered, and which might offer an alternative perspective on why orthographic effects are so pervasive. Language is often used to refer to things in the world. There is emerging evidence that as a spoken word unfolds, visually based perceptual information is rapidly activated, and that such information directs a listener's attention to potential referents in the world as the acoustic signal unfolds (e.g., Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007). Under some circumstances, those referents are printed words, for example, when one is looking for an exit (e.g., *Start watching for the Rochester exit*) or when a child is shown his name in print (e.g., *That's how you write Max*). We propose that orthographic information might serve the same purpose in identifying potential referents as other types of visually based perceptual representations activated during spoken-word recognition. If this perspective is correct, then listeners should (a) activate orthographic representations as they are hearing a word, with the same time course as other semantic and perceptual representations, and (b) use orthographic information to map a spoken word onto its potential printed referents.

We propose that orthographic information becomes activated during the earliest moments of spoken-word recognition and that listeners use this information to map the spoken word onto potential printed referents, just as they use perceptual representations to map words onto objects and depicted objects. We contrast this with the hypothesis that objects, pictures, and words activate names that mediate the mapping between a spoken word and its potential referents. Dahan and Tanenhaus (2005) provided evidence against the phonological (implicit naming) hypothesis for mapping spoken words onto pictures. They showed that upon hearing a spoken word, listeners treat an object that is visually similar to the shape associated with the spoken word as a potential referent even though that object's name does not match the spoken word (e.g., looking at a picture of a rope upon hearing the word *snake*). However, the phonological hypothesis is more plausible for printed words than it is for pictures because visual word recognition rapidly makes available phonological representations.

The orthographic hypothesis makes two strong predictions that distinguish it from the phonological hypothesis. The first is that orthographic effects should emerge during the earliest moments of spoken-word recognition. The second is that in the absence of sufficient preview time to read the printed words, orthographic rather than phonological similarity will mediate the mapping between an unfolding spoken word and potential printed referents.

In order to evaluate these predictions, we used a modified version of the visual-world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This paradigm has been used extensively in studies of spoken-word recognition beginning with Allopenna, Magnuson, and Tanenhaus (1998), who examined listeners' eye movements to pictures in a visual display upon hearing the name of one of the pictures. A large body of work with this paradigm has shown that fixations to potential referents provide valuable information about the time course of lexical competition in spoken-word recognition, including the use of fine-grained subphonemic acoustic information (e.g., McMurray, Clayards, Tanenhaus, & Aslin, 2008; McMurray, Tanenhaus, & Aslin, 2002; Salverda, Dahan, & McQueen, 2003).

We used a version of this paradigm with printed words, introduced by McQueen and Viebahn (2007), to evaluate the time course of activation of orthographic information in spoken-word recognition in conjunction with the use of that information to identify a printed referent. McQueen and Viebahn repeated Allopenna et al.'s (1998) experiment using printed words instead of pictures, finding strikingly similar results. On the basis of these results, some have assumed that the mapping between spoken word and printed word is mediated by phonological representations. For instance, Huettig and McQueen (2007) used the results of several printed-words experiments to make arguments about the time course with which different types of information (including phonological information) are accessed and used in mapping spoken words onto referents in visual-world studies. Huettig and McQueen's results (and those of McQueen & Viebahn, 2007) are, however, fully consistent with our hypothesis that listeners use the orthographic information that is activated as a consequence of hearing a spoken word to map spoken words onto printed words. Although the sound-form associated with the printed referent matches the phonological representation of the spoken target word, the visual word-form of a printed referent also perfectly matches the orthographic representation associated with the spoken target word. Thus, even though McQueen and Viebahn found that listeners rapidly fixated printed referents of spoken words, and that the time course of fixations to the referents was very similar to the time course of fixations to pictures in studies with picture referents (e.g., Allopenna et al., 1998), their study does not establish whether the representations that are used to map spoken words onto printed words are phonological or orthographic.

In order to examine whether orthographic representations mediate the mapping of spoken words onto printed words, one needs to examine looks to printed competitors that vary in their degree of phonological (but not orthographic) overlap with the target and looks to printed competitors that vary in their degree of orthographic (but not phonological) overlap with the target. If the search for the printed referent of a spoken word is mediated by an orthographic representation that is rapidly available upon hearing the spoken word, fixations to printed competitors should be sensitive to the degree of orthographic overlap, but not to the degree of phonological overlap between the target and competitor. If, on the other hand, only phonological representations mediate the matching process, then fixations to printed competitors should be primarily sensitive to phonological overlap and not to orthographic overlap.

In the experiments presented in this article, participants saw four words on a computer screen while they heard a sentence that referred to one of the words (e.g., *Click on the word bead*). The display included the target word, a competitor, and two unrelated distractors. By varying the degree of orthographic and phonological overlap between the target and competitor, we were able to examine whether orthographic information contributed to the identification of the printed referent. In Experiment 1, the competitors had the same degree of orthographic overlap with the target and we compared looks to a competitor with a lower degree of phonological overlap with the target (e.g., target *bead*, competitor *bear*) to looks to a competitor with a higher degree of phonological overlap with the target (e.g., target *bead*, competitor *bean*). If phonological representations mediate looks to the printed words, we should see more looks to competitor words compared to unrelated words as well as a difference in the proportion of looks to higher and lower phonological-overlap competitors. In Experiment 2, we used homophone competitors to hold phonological similarity constant while manipulating the degree of orthographic overlap with the target. We compared looks to a competitor with a lower degree of orthographic overlap with the target (e.g., target *bead*, competitor *bare*) to looks to a competitor with a higher degree of orthographic overlap with the target (e.g., target *bead*, competitor *bear*). Here, the orthographic hypothesis predicts stronger competitor effects for the higher overlap condition, with the timing of the effects tightly linked to the processing of the vowel in the target word.

Experiment 1

Method

Participants. Twenty students at the University of Rochester, all native speakers of English, were paid for their participation in the experiment. They reported normal or corrected-to-normal vision and normal hearing.

Materials. We constructed 16 triples. Each triple consisted of a target word (e.g., *bead*), a competitor with a high degree of phonological overlap (e.g., *bean*), and a competitor with a lower degree of phonological overlap (e.g., *bear*). Target words were nonhomophonous. Importantly, each type of competitor had the same degree of orthographic overlap with the target. Each triple was associated with two orthographically and phonologically unrelated distractors.[1] Across conditions, higher overlap and lower overlap competitors were matched for frequency. Word frequencies were estimated using the frequency counts reported in Francis and Kučera (1982). The mean frequency of the competitor was 32 per million in the higher overlap condition and 31 per million in the lower overlap condition. Distractors were matched for frequency with the target. The mean frequency of the target was 96 per million; the mean frequency of the distractor was also 96 per million. The 16 experimental stimulus sets are listed in Appendix A. In addition, 46 filler sets were constructed, each consisting of four words. Sixteen of these filler sets were included to discourage participants from developing expectations that in displays with two orthographically similar words, those words were likely targets. In half of these filler sets the distractors had a lower degree of phonological overlap with each other, and in the other half the distractors had a higher degree of phonological overlap with each other. The remaining 30 filler sets consisted of four words with no orthographic or phonological overlap. Ten of these sets were selected as practice trials and appeared at the beginning of the experiment, in order to help participants familiarize themselves with the experimental task and procedure.

Each target word was recorded in a fixed instruction sentence that referred to a printed word, for example, *Click on the word bead*. The target word was preceded by a brief prosodic break. Sentences were recorded digitally, in a randomized order, by a male native speaker of English who was unaware of the goal of the study. The sentences were labeled and edited using a speech editor. The mean duration of the target word was 563 ms (SD = 113).

Design. Each trial consisted of the presentation of the four words from a stimulus set along with an instruction sentence. A set of 10 practice trials was presented at the beginning of the experiment. Two lists were constructed by varying whether the competitor had a lower or higher degree of phonological overlap with the target. Within each list, half of the experimental trials had a lower overlap competitor, and half of the experimental trials had a higher overlap competitor. Four random orders were created for each list, with the constraint that there were no more than two consecutive experimental trials, and with the position of the words in the display randomized for each trial. Ten participants were randomly assigned to each list, and an approximately equal number of participants was assigned to each randomization.

Procedure. Eye movements were monitored using a head-mounted SR EyeLink II eye-tracking system (www.sr-research.com), sampling at 250Hz. The eye tracker was calibrated prior to the experiment. Each trial started with the presentation of a small dot in the center of the screen, which the participant fixated and clicked on to perform a drift correction. Subsequently, a spoken instruction was presented through headphones. The visual display appeared 200 ms before the onset of the target word in the sentence (see Figure 1), following the procedure used by McQueen and Viebahn (2007) and Huettig and McQueen (2007). This was done in order to prevent listeners from strategically reading the words and, therefore, accessing the words' phonological codes prior to hearing the target word. Words were presented in lowercase Times New Roman font. Each word was approximately 3 to 4° wide and its center appeared approximately 8° from the center of the screen. Participants used a computer mouse to click on the target word.

Fixations were coded automatically using custom-made software developed by Edward Longhurst. Fixations were coded as

[1] A few of the distractors had a small degree of orthographic overlap with the target. However, if those distractors attracted fixations, this should result only in reduced competition effects, regardless of experimental condition.

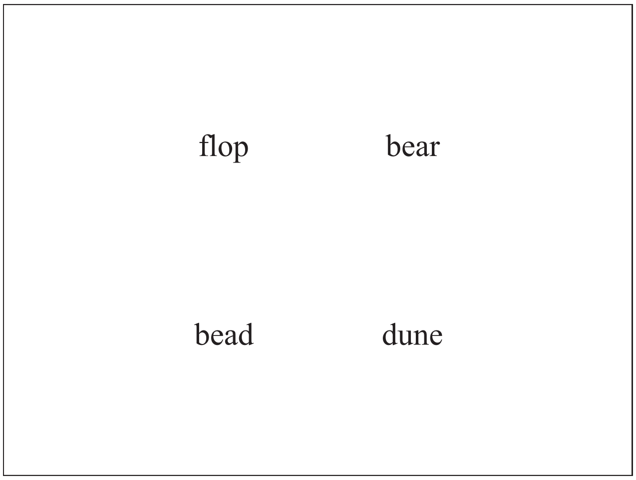

Figure 1. Example of a visual display.

directed to the target, the competitor, either of the distractors, or another location on the screen. Fixations falling within a square region of 4 × 4° from the center of a word were coded as a fixation on that word.

Results

One trial (0.3% of the data) was excluded from the analyses because the participant clicked on the competitor instead of the target.

We computed, for each 4-ms time bin, the proportion of fixations to each type of word in the visual display, for each participant and for each item. Figure 2 presents the mean proportion of fixations to the target, the competitor, and the averaged distractors in the lower overlap condition and the higher overlap condition, beginning at the onset of the target word. At word onset, fixation proportions to each type of word did not differ. Beginning at about 200 ms, fixations to the target and competitor started to increase, whereas fixations to the distractors decreased. Taking into account an estimate of 150–200 ms for programming and executing an eye movement (Hallett, 1986), fixations to printed referents were closely time locked to relevant information in the speech signal. Fixations to the competitor exceeded fixations to the distractors until they converged at around 1,000 ms after target word onset.

We first examined competition effects in each experimental condition. In order to test whether the difference in fixations to the competitor versus the distractors was significant, we computed, for each experimental condition, the ratio between the proportion of fixations to the competitor and the sum of the proportion of fixations to the competitor and the averaged distractor, following a procedure introduced by Salverda et al. (2007). A ratio of .5 would indicate that the competitor did not attract more fixations than the distractors. The competitor ratio was computed for the time interval extending from 200 ms after the onset of the target word to 200 ms after the offset of the target word (i.e., 200 to 763 ms). In order to avoid problems associated with performing some statistical tests on proportion data, we submitted ratio and proportion data to an empirical logit transformation prior to performing statistical tests (see Jaeger, 2008). For reasons of clarity, we present the original ratio or proportion values in the text. One-sample t tests showed that the competitor ratio was significantly different from .5 in the lower overlap condition (ratio = .72), $t_1(19) = 5.66$, $p < .001$; $t_2(15) = 5.67$, $p < .001$, as well as in the higher overlap condition (ratio = .69), $t_1(19) = 4.06$, $p < .001$; $t_2(15) = 5.25$, $p < .001$. This demonstrates that participants were more likely to fixate the competitor than the distractors in both experimental conditions.

We tested the effect of experimental condition by computing, for each participant and each item, the mean proportion of fixations to the competitor during the time interval from 200 ms after the onset of the target word to 200 ms after the offset of the target word (i.e., during a time interval from 200 to 763 ms after word onset). During this interval, listeners were equally likely to fixate a competitor with a higher degree of phonological overlap with the target (22%) compared to a competitor with a lower degree of phonological overlap with the target (22%). Planned t tests revealed that this difference was nonsignificant, $t_1(19) = 0.15$, $p = .442$; $t_2(15) = 0.35$, $p = .367$.

The degree of phonological overlap between target and competitor did not affect the likelihood of fixating the competitor, suggesting that fixations to printed words are not mediated by phonological information. However, the results are compatible with the hypothesis that fixations to printed words are mediated by orthographic information. Competitors in each overlap condition were matched for orthographic overlap with the target and had greater overlap with the target than did the unrelated distractors, and the results show clear and equivalent-sized competitor effects for the two overlap conditions.

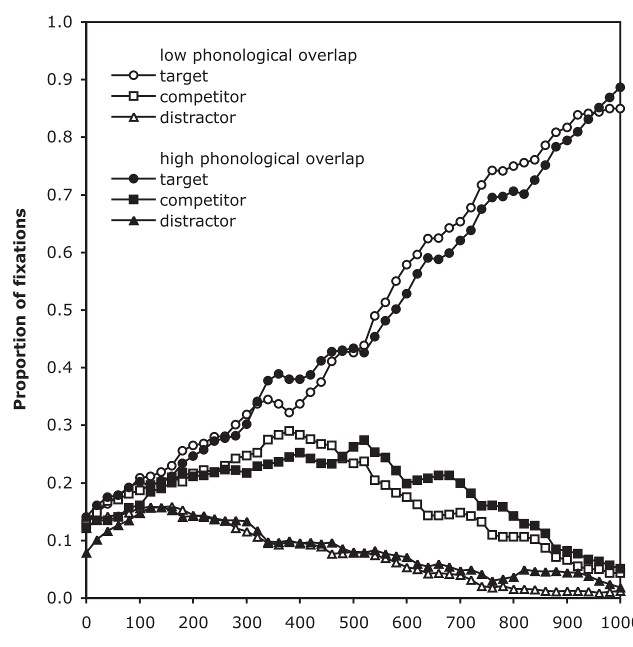

Figure 2. Proportion of fixations to the target, the competitor, and the averaged distractor in the lower phonological-overlap condition and the higher phonological-overlap condition in Experiment 1.

Experiment 2

The results of Experiment 1 suggest that fixations to printed words upon hearing a spoken word are mediated by orthographic information activated upon hearing the spoken word. However, in Experiment 1, the degree of orthographic overlap between target and competitor was identical between experimental conditions (e.g., target *bead*, competitor *bean*, in the higher phonological-overlap condition, or competitor *bear*, in the lower phonological-overlap condition). Therefore, the results of Experiment 1 do not provide detailed information about the time course of the activation of orthographic information. Moreover, the absence of a phonological effect is, of course, a null result.

In Experiment 2, we varied the degree of orthographic overlap between target and competitor and examined how this affected the probability of fixating the competitor. Importantly, we made sure that the phonological overlap between the target and each type of competitor was identical by using competitor words that were homophonous (e.g., target *bead*, competitors *bear* and *bare*). Each competitor had a low degree of phonological overlap with the target (cf. /bid/, /bɛr/, /bɛr/). If fixations to printed words are mediated by orthographic information activated upon hearing the spoken target word, we would expect to see more fixations to the competitor in the higher overlap condition (e.g., target *bead*, competitor *bear*) than in the lower overlap condition (e.g., target *bead*, competitor *bare*). Importantly, differences in competitor fixations as a function of degree of orthographic overlap between target and competitor would provide important information concerning the time course of activation of orthographic information in spoken-word recognition.

Method

Participants. Twenty students at the University of Rochester, all native speakers of English, were paid for their participation in the experiment. They reported normal or corrected-to-normal vision and normal hearing.

Materials. Sixteen triples were constructed. Each triple consisted of a target word (e.g., *bead*), a competitor with a high degree of orthographic overlap with the target (e.g., *bear*) and a homophonous competitor with a lower degree of orthographic overlap with the target (e.g., *bare*). Target words were nonhomophonous. Each triple was associated with two orthographically and phonologically unrelated distractors.[2] Across conditions, the lower overlap and higher overlap competitors were matched for frequency, using frequency counts reported in Francis and Kučera (1982). The mean frequency of the competitor was 29 per million in the lower overlap condition and 28 per million in the higher overlap condition. Distractors were matched for frequency with the target. The mean frequency of the target was 184 per million; the mean frequency of the distractor was 182 per million. The 16 experimental stimulus sets are listed in Appendix B. In addition, 46 filler sets were constructed, each consisting of four words. Sixteen of these filler sets were included to discourage participants from developing expectations that in displays with two orthographically similar words, those words were likely targets. Half of these filler sets had two distractors with a low degree of orthographic overlap, and the other half had two distractors with a high degree of orthographic overlap. The remaining 30 filler sets consisted of four words with no orthographic or phonological overlap. Ten of these sets were selected as practice trials and appeared at the beginning of the experiment in order to help participants familiarize themselves with the experimental task and procedure.

Each target word was recorded in a fixed instruction sentence that referred to a printed word, for example, *Click on the word bead*. The target word was preceded by a brief prosodic break. Sentences were recorded digitally, in a randomized order, by the same speaker who recorded the materials for Experiment 1. The sentences were labeled and edited using a speech editor. The mean duration of the target word was 517 ms ($SD = 68$).

Design and procedure. The experimental design and procedure were identical to those of Experiment 1.

Results

Two trials (0.6% of the data) were excluded from the analyses because the participant clicked on the competitor instead of the target.

We computed, for each 4-ms time bin, the proportion of fixations to each type of word in the visual display, for each participant and for each item. Figure 3 presents the mean proportion of fixations to the target, the competitor, and the averaged distractors in the lower overlap condition and the higher overlap condition, starting at the onset of the target word. At word onset, fixation proportions to each type of word did not differ. From about 200 ms on, fixations to the target and competitor started to increase, whereas fixations to the distractors decreased. Fixations to the competitor exceeded fixations to the distractors until they converged at around 1,000 ms after target word onset. Importantly, the proportion of fixations to the competitor increased faster and reached a higher peak in the higher overlap condition than in the lower overlap condition.

We first examined competition effects in each experimental condition. In order to test whether the difference in fixations to the competitor and the distractors was significant we computed, for each experimental condition, the ratio between the proportion of fixations to the competitor and the sum of the proportion of fixations to the competitor and the averaged distractor. This competitor ratio was computed for the time interval extending from 200 ms after the onset of the target word to 200 ms after the offset of the target word (i.e., 200 to 717 ms, given that the mean duration of the target word was 517 ms). One-sample t tests showed that the competitor ratio was significantly different from .5 in the lower overlap condition (ratio = .64), $t_1(19) = 2.24$, $p < .05$; $t_2(15) = 2.08$, $p < .05$, as well as in the higher overlap condition (ratio = .71), $t_1(19) = 5.66$, $p < .001$; $t_2(15) = 5.62$, $p < .001$. Thus, participants were more likely to fixate the competitor than the distractors in both experimental conditions.

Because the lower and higher orthographic-overlap competitors had the same degree of phonological overlap with the target, comparing the two conditions allows us to focus on

[2] A few of the distractors had a small degree of orthographic overlap with the target. However, if those distractors attracted fixations, this should result only in reduced competition effects, regardless of experimental condition.

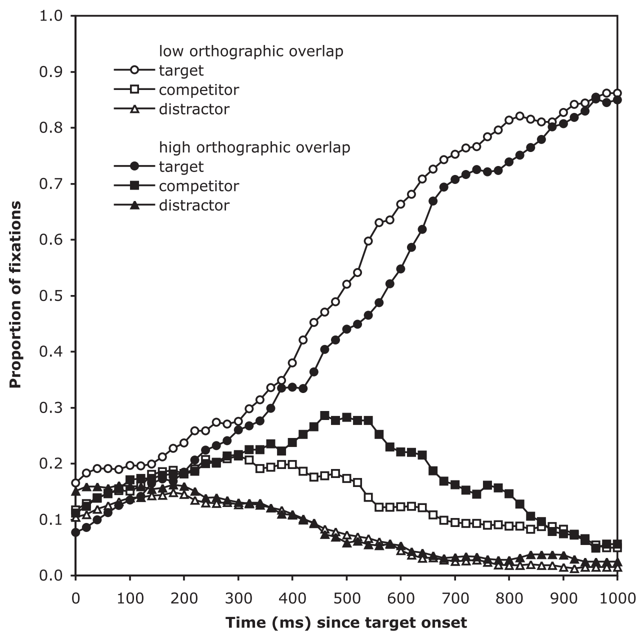

Figure 3. Proportion of fixations to the target, the competitor, and the averaged distractor in the lower orthographic-overlap condition and the higher orthographic-overlap condition in Experiment 2.

effects due to activation of orthographic information—and to examine in detail the time course of the activation of such information. We tested the effect of experimental condition by computing, for each participant and each item, the mean proportion of fixations to the competitor between 200 and 717 ms after word onset (i.e., during the acoustic lifetime of the target word, assuming a 200-ms delay for programming an eye movement). During this interval, listeners were more likely to fixate a competitor with a higher degree of orthographic overlap with the target (23%) compared to a competitor with a lower degree of orthographic overlap with the target (17%). Planned t tests showed that this difference was significant, $t_1(19) = 2.42$, $p < .05$; $t_2(15) = 3.08$, $p < .005$. This result suggests that fixations to printed words were mediated by orthographic information associated with the spoken target word. The results also show that such information was activated during the acoustic lifetime of the target word.

Figure 4 focuses on the proportion of fixations to lower overlap competitors and higher overlap competitors. This allows us to examine in more detail the effect of orthographic overlap between target and competitor on fixations to the competitor. Until about 300 ms after target word onset, participants are equally likely to fixate the competitor as a function of experimental condition. Shortly after 300 ms, fixations to higher overlap competitors begin to exceed fixations to lower overlap competitors. This difference in competitor fixations between conditions increases rapidly and reaches its peak around 500 ms. When evaluating these results, it is important to keep in mind that the initial sound(s) of the target word, preceding the vowel, were fully consistent with each competitor. The mean duration of this part of the target word was 103 ms ($SD = 67$). Taking into account a 200-ms delay for programming an eye movement, the earliest point at which we would expect to see the emergence of a difference in competitor fixations between conditions mediated by orthographic information is therefore around 300 ms after the onset of the target word. And this is precisely what was observed (see Figure 4). Moreover, the emergence of a difference in competitor fixations between experimental conditions was closely time locked to the vowel of the target word. When we restrict our analysis to the time window from 303 to 511 ms after the onset of the target word (i.e., the time interval corresponding to the presentation of the vowel, which was 208 ms in duration on average [$SD = 57$], taking into account a 200-ms delay for programming an eye movement), planned t tests show that the effect of condition approaches significance by participants, $t_1(19) = 1.52$, $p = .073$, and remains significant by items, $t_2(15) = 2.09$, $p < .05$.

The results of Experiment 2 show that the degree to which listeners fixate a printed word upon hearing a spoken word is affected by the degree of orthographic overlap between spoken word and printed word. This suggests that the mapping between the speech signal and printed words is mediated by orthographic information activated upon hearing the spoken target word. Before accepting this conclusion, however, it is important to evaluate an alternative explanation of our results based on phonological representations. The absence of a phonological effect in Experiment 1 rules out any possible effects of lexically mediated phonological representations. If there were any such effects, they would have been reflected in more looks to higher phonological-overlap competitors. However, one might still argue that the effects observed in

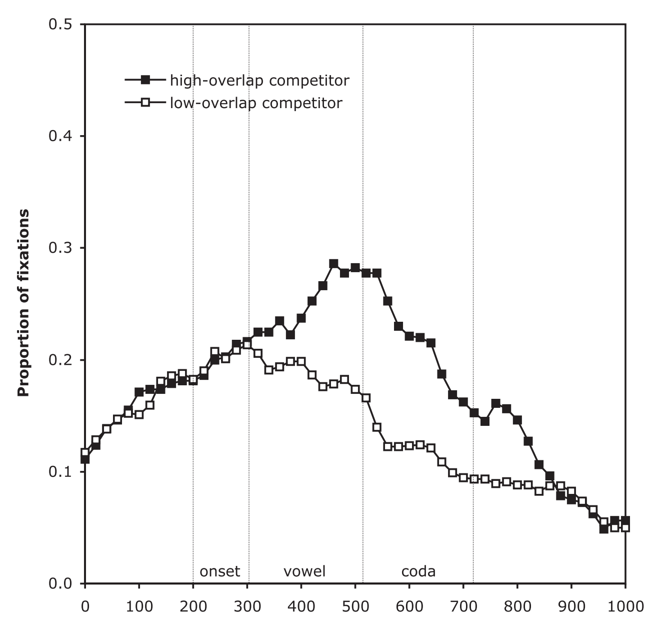

Figure 4. Proportion of fixations to the competitor in the lower orthographic-overlap condition and the higher orthographic-overlap condition in Experiment 2. Vertical lines relate eye fixations to information in the target word, averaged across items and assuming a 200-ms delay for programming an eye movement.

Experiment 2 are mediated by sublexical phonological representations activated by the printed words.[3]

The printed-words based, sublexical phonological activation account assumes that listeners rapidly access sublexical phonological information upon seeing a printed word and that these phonological representations are mapped onto a phonological representation of the spoken target word. The idea is that in the higher overlap condition, participants see the competitor *bear* and activate the sublexical phonological representation /i/ because in many English words, the letter sequence "ea" corresponds to the vowel /i/. The phonological representation /i/ matches the vowel in the spoken target word *bead*, and this is the reason that participants are likely to fixate the higher overlap competitor. In contrast, the phonological representation /i/ is not activated upon seeing (letter sequences in) the lower overlap competitor *bare*, and sublexical phonological information activated upon seeing this competitor would thus match the spoken target word to a lesser extent.

We find it implausible that sublexical phonological representations would have strong effects when there is no suggestion of a lexical phonological effect in Experiment 1. Nonetheless, the sublexical phonological activation account is difficult to rule out in principle. However, the design of our experiment and the timing of the presentation of the visual display allowed us to conduct analyses that evaluate the plausibility of this account. In particular, because the spoken word is presented exactly 200 ms after the appearance of the visual display, there is time for only a single fixation to a printed word prior to the onset of the spoken word. Results from the reading literature suggest that phonological information can be accessed for the word currently fixated and, at least to a certain extent, for the following word (Rayner, 1998), with parafoveal phonological effects extending to a word several degrees to the right of fixation but only several characters (typically less than .5°) to the left of fixation. If sublexical phonological information is accessed and used to map the spoken target word onto the printed words in the visual display, we would expect that the effect of condition observed in Experiment 2 would be driven by the subset of trials in which the participant's initial fixation was to the competitor or the competitor was to the right of fixation. We performed a contingent analysis by excluding such trials from the data (88 trials in total, distributed equally across conditions; 27.7% of the data). We analyzed the remaining trials (see Figure 5).

The results show an effect of condition that is remarkably similar in timing, magnitude, and duration to the effect found for the complete data set (cf. Figure 4). For the reduced data set, during the time interval corresponding to the presentation of the target word, that is between 200 and 717 ms after word onset, the mean proportion of fixations to the competitor was 11% in the lower overlap condition and 17% in the higher overlap condition. Planned t tests confirmed that this difference was significant, $t_1(19) = 2.58$, $p < .01$; $t_2(15) = 2.62$, $p < .01$. During the time interval corresponding to the presentation of the vowel of the target word, that is, between 303 and 511 ms after word onset, the mean proportion of competitor fixations was 13% in the lower overlap condition and 19% in the higher overlap condition. Planned t tests showed that this difference was also significant, $t_1(19) = 1.83$, $p < .05$; $t_2(15) = 2.83$, $p < .01$.

The explanation based on sublexical phonological representations assumes that these representations are available as the spoken word unfolds. However, the contingent analyses of the results of Experiment 2 show that the effect of orthographic similarity was not contingent on having access to the precise word form of the competitor prior to hearing the spoken word. This finding lends further support to our interpretation of the results as revealing the time course of the activation of orthographic information in spoken-word recognition.

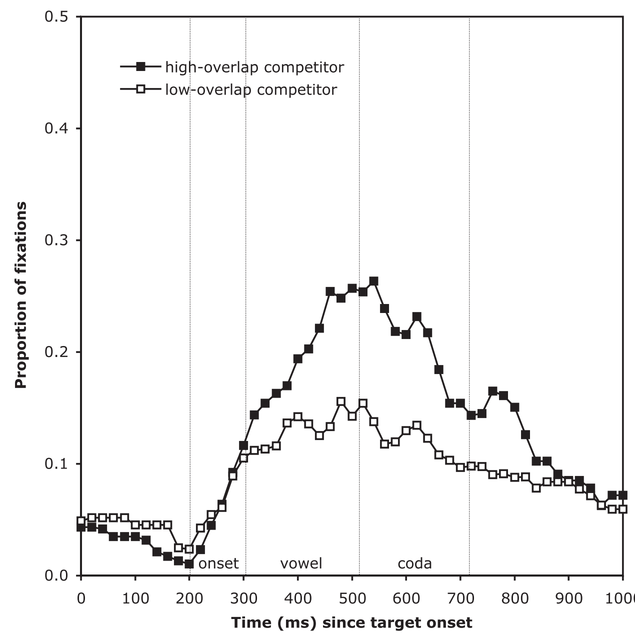

Figure 5. Proportion of fixations to the competitor in the lower orthographic-overlap condition and the higher orthographic-overlap condition in Experiment 2, for the subset of trials where at 200 ms after the onset of the target word the competitor was not fixated and not located to the right of fixation. Vertical lines relate eye fixations to information in the target word, averaged across items and assuming a 200-ms delay for programming an eye movement.

One might argue that when the word is eccentric to the locus of fixation, information about the orthographic form of the word would also be less available, and thus any orthographic effects should be reduced. However, although effects of eccentricity might reduce the overall strength of competitor effects, they should not interact with the magnitude of the orthographic differences between conditions, which is exactly what we find. To see why this is the case, assume that the spoken word makes available an orthographic representation and initiates visual search for the referent. The likelihood that a shift in attention to a word will result in a subsequent fixation to that word would depend upon the goodness of the match of the printed word to that representation. Thus orthographic effects on the likelihood of a fixation do not depend on having a detailed orthographic representation of a competitor prior to an attentional shift. In sum, the additional analyses do not support the sublexical phonological activation account, whereas they are consistent with the orthographic hypothesis.

[3] We thank an anonymous reviewer for suggesting this explanation.

General Discussion

By varying the degree of orthographic overlap between a printed referent of a spoken word and its competitor, we were able to track the time course of the activation of orthographic information during the processing of a spoken word. Our results establish that in a version of the printed-words paradigm with brief preview of the visual display, participants use orthographic information to identify the printed referent of a spoken word. Fixations to printed words upon hearing a spoken target word were found to be sensitive to the degree of orthographic overlap between target and competitor (Experiment 2) but not to the degree of phonological overlap between target and competitor (Experiment 1). The results of Experiment 2 further provide detailed information about the time course of the activation of orthographic information in spoken-word recognition. These results provide compelling evidence that orthographic information is activated very rapidly, from the earliest moments in time, as a spoken word is heard and processed. It appears that there is virtually no delay between the processing of phonological information in the speech signal and the activation of associated orthographic representations. Our findings thus complement and extend the results of ERP studies by Perre and Ziegler (2008) and Pattamadilok et al. (2009), which first demonstrated that orthographic information is activated rapidly and relatively early in the recognition process.

Our results parallel those of a visual-world study by Dahan and Tanenhaus (2005). In their study, participants rapidly fixated a picture of a snake upon hearing the spoken word *snake*. But they also frequently made transitory fixations to a visual competitor of the target word (in this case, a coiled rope), even if they had already looked at the competitor during preview. This result suggests that upon hearing the word, a listener activates a visual representation that draws the listener's attention to visually similar potential referents. Thus, the mapping between the speech signal and pictures in a visual display is mediated, at least in part, by the rapid activation of visually based conceptual representations associated with a spoken word. Our results demonstrate that visually based orthographic representations (e.g., those associated with the visual word-form *snake*) are also activated very rapidly upon hearing a spoken word and that listeners can rapidly use those representations to identify a printed referent.

The vast majority of studies that examined the activation of orthographic information in spoken-word recognition are concerned with demonstrating that orthographic information is activated automatically, that is, nonstrategically, upon hearing a spoken word. The results of these studies are compelling and unequivocal (see e.g., Taft et al., 2008, and Pattamadilok et al., 2009, for recent evidence). The current study was not designed to contribute to the literature on this topic. Nevertheless, the rapid influence of orthographic representations on eye movements to printed words reveals that spoken words can be mapped extremely quickly onto printed referents. This finding is fully consistent with the results of the large body of work showing that orthographic information is automatically accessed as a consequence of hearing a spoken word.

Why are visual representations used to map spoken words onto objects or depicted referents in the visual world, and why—at least with short preview—are orthographic representations used to map spoken words onto printed words?[4] It might seem that linguistic encoding of printed words would be more efficient. In fact, we believe that there are circumstances where listeners might strategically use linguistic coding. In general, however, representations of the visual world serve other functions than to be mapped onto linguistic representations. Moreover, linguistic recoding is inefficient for visual search because there is nothing in a linguistic representation that directly links information to its associated spatial location. In contrast, the perceptual information accessed when a word is heard can be used either to map that word onto its referent or to build an internal representation. Thus, it makes functional sense for linguistic coding of the visual world to be under strategic control, whereas the activation of visually based perceptual representations associated with linguistic processing would be a more automatic process.

Our results have clear implications for our understanding of both the access and use of orthographic representations in spoken-word recognition and for our understanding of the linking hypothesis in the printed-words version of the visual-world paradigm. Most generally, they provide further support for the notion of a "distributed lexicon" (Elman, 2004) in which features that co-occur with a word form may be incorporated into a less focused representation of meaning. These properties are correlated with one another and will upon presentation coactivate each other. In contrast to a more restricted view of the lexicon in which only linguistically relevant features are stored as part of a "lexical entry," the distributed-lexicon hypothesis maintains that processing of a spoken word activates a rich set of interconnected representations, including the word's phonological form, combinatory linguistic information such as verb argument structure (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus, Carlson, & Trueswell, 1989), indexical information (Creel, Aslin, & Tanenhaus, 2008; Goldinger, 1998), conceptual/perceptual knowledge (Barsalou, 1999; Dahan & Tanenhaus, 2005; Fischer & Zwaan, 2008; Martin & Chao, 2001; Pirog-Revill, Aslin, Tanenhaus, & Bavelier, 2008), and motor representations (Hauk, Johnsrude, & Pulvermüller, 2004; Tyler & Moss, 2001).

Some types of perceptual/motor representations might play an integral role in understanding the meaning of sentences (Glenberg, 2007; Richardson, Spivey, Barsalou, & McRae, 2003). In addition, some of these representations are central to one of the most basic functions of spoken language, which is to use words to refer to events and entities in the world.

[4] With longer preview of printed words, we expect that participants will typically have fixated multiple words prior to the presentation of the spoken word and that phonological effects should emerge as a by-product of the participant reading the words. In pilot work we replicated Experiment 1 with longer preview. The time between the presentation of the display and the spoken word ranged from 1,000–1,500 ms. Here we found an effect of degree of phonological overlap on competitor fixations.

References

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

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–609.

Chereau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition, 102,* 341–360.

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.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition, 106,* 633–664.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken word recognition. *Psychonomic Bulletin & Review, 12,* 453–459.

Donnenwerth-Nolan, S., Tanenhaus, M. K., & Seidenberg, M. S. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 170–180.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences, 8,* 301–306.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology, 61,* 825–850.

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

Frost, R., & Ziegler, J. C. (2007). Speech and spelling interaction: The interdependence of visual and auditory word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 107–118). Oxford, England: Oxford University Press.

Glenberg, A. M. (2007). Language and action: Creating sensible combinations of ideas. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 361–370). Oxford, England: Oxford University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York, NY: Wiley.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111,* 662–720.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41,* 301–307.

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15,* 985–1018.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57,* 460–482.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformed or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology, 11,* 194–201.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review, 15,* 1064–1071.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* B33–B42.

McQueen, J. M., & Viebahn, M. C. (2007). Tracking the recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60,* 661–671.

Pattamadilok, C., Perre, L., Dufau, S., & Ziegler, J. C. (2009). On-line orthographic influences on spoken language in a semantic task. *Journal of Cognitive Neuroscience, 21,* 169–179.

Peereman, R., Dufour, S., & Burt, J. S. (2009). Orthographic influences in spoken word recognition: The consistency effect in semantic and gender categorization tasks. *Psychonomic Bulletin & Review, 16,* 363–368.

Perre, L., Midgley, K., & Ziegler, J. C. (2009). When beef primes reef more than leaf: Orthographic information affects phonological priming in spoken word recognition. *Psychophysiology, 46,* 739–746.

Perre, L., & Ziegler, J. C. (2008). On-line activation of orthography in spoken-word recognition. *Brain Research, 1188,* 132–138.

Pirog-Revill, K., Aslin, R. N., Tanenhaus, M. K., & Bavelier, D. (2008). Neural correlates of partial lexical activation. *Proceedings of the National Academy of Sciences, USA, 105,* 13111–13115.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27,* 767–780.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51–89.

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated subphonetic variation on lexical competition. *Cognition, 105,* 466–476.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 546–554.

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language, 58,* 366–379.

Tanenhaus, M. K., Carlson, G. N., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes, 4,* 211–234.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8,* 513–520.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995, June 16). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences, 5,* 244–252.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5,* 683–689.

(*Appendices follow*)

Appendix A

Stimulus Sets for Experiment 1

	Competitor			
Target	Higher overlap	Lower overlap	Distractor	Distractor
bead (1)	bean (5)	bear (57)	flop (1)	dune (1)
clap (1)	clan (2)	claw (1)	moth (1)	dire (1)
earn (16)	earl (12)	ears (37)	kick (16)	spun (16)
face (371)	fade (2)	fare (7)	best (351)	city (393)
flag (16)	flat (67)	flaw (3)	bulk (16)	jeep (16)
fool (37)	food (147)	foot (70)	wash (37)	hurt (37)
mate (21)	maze (6)	mare (16)	loop (21)	cope (21)
nose (60)	node (2)	none (108)	text (60)	duty (61)
paid (145)	pain (88)	pair (50)	sent (145)	club (145)
pant (10)	pals (1)	park (94)	slug (10)	wipe (10)
peas (24)	peak (16)	pear (6)	slid (24)	jump (24)
poke (1)	pose (11)	pore (2)	fawn (1)	jest (1)
span (19)	spat (9)	spar (3)	whip (19)	fury (19)
steam (17)	steal (5)	steak (10)	chaos (17)	dandy (17)
stem (29)	step (131)	stew (5)	cure (28)	dull (27)
work (760)	worm (4)	worn (23)	make (794)	here (750)

Note. Competitor conditions varied in the degree of phonological overlap with the target. Lexical frequencies are in parentheses, as reported in Francis and Kučera (1982).

Appendix B

Stimulus Sets for Experiment 2

	Competitor			
Target	Higher overlap	Lower overlap	Distractor	Distractor
bead (1)	bear (57)	bare (29)	gild (1)	frog (1)
boat (72)	boar (1)	bore (24)	hill (72)	wine (72)
brash (1)	brake (2)	break (88)	spoke (87)	dance (90)
dead (174)	dear (54)	deer (13)	move (171)	late (179)
dip (6)	die (73)	dye (—)	lad (6)	rub (6)
fail (37)	fair (77)	fare (7)	mood (37)	crew (36)
half (275)	hale (2)	hail (10)	week (275)	seen (279)
ham (19)	hay (19)	hey (15)	lid (19)	eve (19)
malt (1)	male (37)	mail (47)	ship (1)	cube (1)
paid (145)	pair (50)	pare (2)	blue (143)	meet (148)
said (1961)	sail (12)	sale (44)	what (1908)	them (1789)
sink (23)	sine (4)	sign (94)	lock (23)	aunt (22)
sold (47)	sole (18)	soul (47)	wave (46)	pink (48)
stark (7)	stare (14)	stair (2)	cough (7)	greet (7)
steam (17)	steak (10)	stake (20)	bunch (17)	graph (17)
talk (154)	tale (21)	tail (24)	club (145)	sort (164)

Note. Competitor conditions varied in the degree of orthographic overlap with the target. Lexical frequencies are in parentheses, as reported in Francis and Kučera (1982).

Received February 18, 2009
Revision received March 30, 2010
Accepted March 31, 2010 ■