Journal of Experimental Psychology:
Human Perception and Performance
2009, Vol. 35, No. 5, 1585–1597

0096-1523/09/$12.00 DOI: 10.1037/a0015870

The Activation of Embedded Words in Spoken Word Identification Is Robust but Constrained: Evidence From the Picture-Word Interference Paradigm

Jeffrey S. Bowers
University of Bristol

Colin J. Davis
Royal Holloway, University of London

Sven L. Mattys, Markus F. Damian, and Derek Hanley
University of Bristol

Three picture-word interference (PWI) experiments assessed the extent to which embedded subset words are activated during the identification of spoken superset words (e.g., *bone* in *trombone*). Participants named aloud pictures (e.g., *brain*) while spoken distractors were presented. In the critical condition, superset distractors contained a semantically related embedded word (e.g., *charm*, which contains *arm*). In Experiment 1, supersets and subsets differed by one phoneme (*charm*/*arm*) and interference effects were obtained when subsets were embedded at the beginning or end of the superset. In Experiment 2, the subsets and supersets differed by three or more phonemes. Interference was obtained for final embedded words aligned with a syllable boundary of the superset (*acrobat*/*bat*) but not otherwise (*pioneer*/*ear*). In Experiment 3, the size of these PWI effects was compared to the effects obtained with the subset words presented in isolation. The implication of these findings for theories of speech perception and production are discussed. We also consider the possible advantages of the PWI task over cross-modal priming and "visual-world" procedures when studying these issues.

Keywords: embedded words, spoken word identification, TRACE, shortlist

All languages are characterized by extensive lexical embedding, that is, short words are embedded within longer words (e.g., *seat* in the spoken word *conceit*). According to one estimate, 83% of spoken polysyllabic words contain at least one embedded word, with embeddings frequent in initial, final, and medial positions (McQueen, Cutler, Briscoe, & Norris, 1995). This raises the question of how the speech perception system correctly identifies the embedding word (the superset) rather than the many embedded words (the subsets) that are also present in the input. In the case of spoken language, this must be achieved despite the fact that word boundaries are poorly marked in the speech stream.

Part of the answer to the speech continuity problem is that listeners exploit a variety of sublexical cues probabilistically associated with word boundaries, for example, metrical stress, phonotactic regularities, and acoustic-phonetic variants (cf., Davis, Marslen-Wilson, & Gaskell, 2002; Mattys, White, & Melhorn, 2005; Salverda, Dahan, & McQueen, 2003). For example, a vast majority of content words in English are stressed initially (Cutler & Carter, 1987) that makes stress a relatively efficient cue for word segmentation. However, sublexical cues are not sufficient. It is generally recognized that lexical knowledge is a major contributor as well. That is, when a string of phonemes is consistent with multiple overlapping lexical forms, multiple lexical entries are coactivated and compete for identification. For example, the word "seat" coactivates the word forms "sea" and "seat" and "eat." Various factors are thought to impact on which of the coactive lexical items are likely to win the competition. Such factors include a bias favoring embedding (i.e., superset) over embedded (i.e., subset) words (McClelland & Elman, 1986; Norris, 1994) and a bias to segment utterances in a way that does not leave any phonemically impossible words stranded, the so-called *possible word constraint* (Norris, McQueen, Cutler, & Butterfield, 1997). A discussion of the relative weightings of various sublexical and lexical cues can be found in Mattys et al. (2005).

Consistent with the lexically driven segmentation hypothesis, a number of studies have reported that form-related words are indeed coactivated during spoken-word recognition, including neighbors and subsets. For example, Shillcock (1990) reported that when listeners are presented with the superset word *trombone*, its subset *bone* is activated as well. As described in the next section, however, the results taken as a whole are somewhat mixed, and whether all types of subsets are activated to the same extent remains unclear.

It is important to resolve these ambiguities given that theories of spoken-word recognition make different predictions about the extent to which embedded words interfere with the recognition of target words. For instance, according to the TRACE model of speech perception (McClelland & Elman, 1986), word-final embedded words are activated, albeit less than word-initial embedded words (see Fig-

---

Jeffrey S. Bowers, Sven L. Mattys, Markus F. Damian, and Derek Hanley, Department of Experimental Psychology, University of Bristol; Colin J. Davis, Department of Psychology, Royal Holloway, University of London.

Correspondence concerning this article should be addressed to Jeffrey S. Bowers, Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, UK. E-mail: J.Bowers@bris.ac.uk

ure 27 in McClelland & Elman, 1986; see also Frauenfelder & Peeters, 1990). By contrast, in the Shortlist model of speech recognition, final-embedded words are not activated (Norris et al., 2006; Figure 1). As in TRACE, subset words in Shortlist receive bottom-up activation (because of form similarity between subset and superset), but the nature of the competition process in Shortlist prevents the bottom-up input from translating into any activation of final-embedded subsets. Therefore, a better characterization of the activation of final embedded words is necessary to distinguish the current implementation of these models. We review these findings below, before introducing a new method to address this issue.

Review of Past Studies That Assessed the Activation of Embedded Words

The most common way to assess the activation of subsets and supersets has been to employ the cross-modal associative priming paradigm. In this task, participants are presented with a spoken superset (either in isolation or in a sentential context) and asked to respond to a written target that is semantically related to the subset or superset. For example, participants might be presented with the spoken sentence: "He carefully placed the trombone on the table," with the written word *rib* presented after *trombone*. The target *rib* is semantically related to the embedded word *bone* within *trombone*, and any facilitation in responding to the visual target *rib* compared to a control visual target is taken as evidence that the subset has been activated—at the level of meaning.

Using this paradigm, Shillcock (1990) assessed priming for two-syllable monomorphemic English words that contained a stressed embedded word in the second syllable. Words were presented in sentences, and participants showed robust priming for visual targets related to the embedded word. Strikingly, priming for the target (e.g., *rib*) was just as large from spoken supersets (e.g., *trombone*) as from the subsets themselves (e.g., *bone*). It is important to note that Shillcock only included supersets with a weak-strong stress pattern, because previously unpublished work failed to obtain priming from strong-weak supersets (Prather & Swinney, 1977). These stress differences might be critical in modulating these word-final subset activation given that the onsets of words in English are generally stressed, and consequently, stress may act as an anchor point for lexical activation (Cutler & Norris, 1988).

A number of additional studies have provided evidence that final embedded words are coactivated, but further constraints on these effects have also been reported. Vroomen and de Gelder (1997) used the cross-modal associative priming paradigm with isolated Dutch words (rather than sentences) and found priming for final embeddings when the subset corresponded to a complete syllable (*boos* in *framboos*, i.e., *angry* and *raspberry*, respectively), but not otherwise (e.g., *wijn* in *zwijn*, i.e., *wine* and *swine*, respectively). This also suggests that the activation of final embedded words depends on the presence of sublexical segmentation cues at the onset of the embedded words; in this case, a syllable boundary. Consistent with this interpretation, Luce and Cluff (1998) reported cross-modal associative priming for both initial and final subsets when the subsets were both strong syllables and words (e.g., as in the spondee *hemlock*). Similarly, Isel and Bacri (1999) observed priming for embedded French words that constituted the final syllable of the superset, consistent with the hypothesis that syllables are efficient cues to lexical segmentation in French (Mehler, Dommergues, Frauenfelder, & Segui, 1981).

Gow and Gordon (1995) assessed priming for words embedded within sentences when the embedded item (e.g., *lips*) was within a word (e.g., *tulips*) as opposed to a stand-alone word (e.g., *two lips*). The authors found priming for a visual target related to *lips* when *two lips* was presented, but not from *tulips*. One possible explanation for why *tulips* failed to support priming (unlike the robust priming obtained with *trombone*, Shillcock, 1990) is that *tulips* is stressed on the first syllable. Indeed, the authors concluded that priming for final embedded words only occurs when the onset of the embedded word is acoustically marked by a stressed syllable.

It should be noted, however, that a few cross-modal studies have failed to observe any priming for final embedded words. In a series of experiments, Norris, Cutler, McQueen, and Butterfield (2006) failed to obtain associative cross-modal priming for subsets embedded in supersets with a weak-strong prosodic structure (e.g., *sedate-time*; most of the items were taken from Shillcock, 1990), and they found inhibitory cross-modal repetition priming for these subsets when embedded in the same supersets (e.g., *sedate-date*). These results suggest that final embedded words are not activated to the level of meaning even when they constitute a stressed syllable. In fact, they indicate that supersets inhibit the lexical-phonological representations of final embedded subsets. Marslen-Wilson, Tyler, Waksler, and Older (1994) assessed cross-modal repetition priming in a task in which the visual target was the same word as the subset (e.g., *trombone* as the spoken superset and *bone* as the visual target). In the case of supersets with final embedded subsets, half had a weak-strong stress pattern, and half a strong-weak stress pattern. No priming was found for either initial or final subsets.

A few studies have employed nonpriming paradigms to investigate the conditions in which embedded items are activated. Cluff and Luce (1990) assessed the perceptual identification of supersets that contained stressed word subsets in both initial and final positions (e.g., *jigsaw*, *latchkey*), with the frequency and neighborhood density of the subsets manipulated so that their identification was fast versus slow when the subsets were presented in isolation. The key finding was that supersets that included easy-to-identify final subsets were better recognized than those that included difficult-to-identify final subsets. Cluff and Luce concluded that final subsets were activated and facilitated the identification of the supersets. By contrast, Luce and Lyons (1999) asked participants to perform lexical decisions or shadow bisyllabic supersets that included word subsets in initial or final syllable position. Supersets with initial subsets were stressed on the first syllable (e.g., *cherish*, which contains *chair*), and supersets with final subsets were, for the most part, stressed on the second syllable (e.g., *conceal*, which contains *seal*). Performance on these items was compared to a set of control words that did not contain a subset (e.g., *flourish*, *concise*). To the extent that subsets were activated and contributed to performance, response times should differ for the superset and control words. The authors reported faster latencies in the lexical decision and shadowing tasks for supersets containing initial subsets but not for supersets containing final subsets. Thus, in this case, there was no evidence that the final subsets were activated, even though the majority of the subsets constituted a stressed syllable.

Recently, the "visual-world paradigm" has been used to address a variety of related issues, including the extent to which initially embedded words are identified to the level of meaning (e.g., Salverda, Dahan, & McQueen, 2003). In this procedure, a participant is first presented with a small number of pictures (typically four), which constitute the visual world for a given trial. After a short delay, a spoken target word is presented and participants' eye movements over the visual display are recorded. The semantic activation generated from the target word is inferred from the participant's eye fixations on the pictures of the display. For instance, if the spoken word *hamster* is presented, an increased proportion of eye fixations to a picture of *ham* compared to an unrelated picture is taken as evidence that *ham* in *hamster* was activated to the level of meaning. In the Salverda et al. (2003) study, the subsets in the spoken supersets were all stress-initial syllables, and they were cross-spliced from either another exemplar of the superset (*hamster*) or from the subset itself (*ham*) recorded in a sentence context. The authors reported robust activation of initial embedded words in both conditions, but they also found more fixations to subset pictures when the subset words were cross-spliced from isolated rather than embedded subsets. This suggests that the degree of semantic activation of subsets is influenced by subphonemic features, such as phoneme duration (the subsets recorded from isolated subsets were longer than the subsets taken from supersets).

To summarize, the extent to which spoken subsets within supersets are activated to the level of meaning has most frequently been assessed using the cross-modal associative priming paradigm, although a few alternative methods have been used as well. A number of these studies provide evidence that embedded subsets are activated to the level of meaning during the process of identifying spoken supersets as long as the subset is marked by a sublexical segmentation cue (e.g., a syllable boundary or a strong syllable). When subsets are syllabically marked, subphonemic features of the subsets (e.g., duration) may also influence the degree to which subsets are activated. However, results have been inconsistent across studies; and, indeed, some studies have failed to obtain any evidence that final embedded items are activated at all, even when their onsets are marked by a stressed syllable. The reason for these discrepant results is not clear.

In the experiments reported below, we assessed the activation of embedded words in a novel way by exploiting the picture-word interference task (e.g., Lupker & Sanders, 1982; Posnansky & Rayner, 1977; Rayner & Springer, 1986). In this task, participants are required to name a picture as quickly as possible while a word (spoken or written) is concurrently presented. Interference manifests itself as slower picture naming times when the distractor word is semantically related to the picture than when it is not, even though participants are instructed to ignore the distractor. For instance, participants are slower to name a picture of a *dog* when the spoken distractor word *cat* is presented compared to the control word *sun*. This is generally considered a form of Stroop interference, and indeed, picture-word interference and classic color-word Stroop are often modeled in the same way (see Roelofs, 2003, for a detailed discussion).

Although this task has been widely used to address theoretical questions in speech production, it can be adapted to address the current issue as well. Indeed, employing the picture-word interference task, Jescheniak and Schriefers (1998) provided evidence consistent with the claim that spoken words activate competitors to the level of semantics. In two studies, the authors asked German speakers to name a picture while a spoken word distractor was presented that overlapped in its initial consonant-vowel segments with a word semantically related to the picture (a synonym). For example, while participants named a picture of a castle (*burg* in German), they were presented with the spoken word distractor *fenchel* (*fennel*), which is form related to a semantic competitor of castle, namely the German word *festung* (a synonym of castle). Stimulus-onset asynchrony (SOA) between picture and word was varied such that the word preceded picture onset (negative SOA), was presented simultaneously with it (SOA = 0 ms) or followed it (positive SOA). The authors reported robust picture-word interference (PWI) effects at –150-, 0-, and +150-ms SOA conditions. Although the authors considered these results in the context of theories of speech production, the findings have implications for speech recognition. Namely, the results suggest that cohorts of spoken words are coactivated, perhaps to the level of semantics.

In the present studies, we used the PWI paradigm to assess the extent to which subsets embedded in spoken supersets are activated to the level of meaning. In one condition, the spoken distractor was semantically related to the picture (the standard condition in which interference is observed), and in the critical conditions, the distractor contained an embedded word in initial or final position that was semantically related to the picture. For example, participants named the picture of a *bus* while hearing the spoken distractor *scar*, which contains the semantically related word *car*. Semantic interference should only be expected from the distractor if the embedded word is activated, first at the level of phonology, and then presumably at the level of meaning (although semantic access may not be required, as noted in the General Discussion). Given that picture-word interference is typically considered an instance of the Stroop phenomenon, any interference is likely to reflect the automatic activation of the embedded word.

To anticipate the main results, we obtained evidence that initial and final subsets are activated when the subset and superset words differ by a single phoneme (e.g., *car* in *scar*; Experiment 1). This occurred even though there were no syllabic cues indicating the onset of the final subsets, contrary to previous findings. We also obtained evidence that final embedded words are activated when the subsets and superset differ by three or more phonemes (Experiment 2), but these effects were restricted to conditions in which the subset constituted (i.e., aligned with) the final syllable of the superset. In addition, the activation of final embedded subsets within the longer supersets was reduced compared to the activation of final embedded subsets within the shorter supersets (Experiment 3). In the General Discussion, we consider the theoretical implications of these findings, and review the possible advantages of the PWI task over other techniques.

Experiment 1

Method

Participants. Thirty-eight undergraduate students from the University of Bristol were recruited and paid a small honorarium for their participation. All were native speakers of British English, had normal or corrected-to-normal eyesight, and had no self-reported auditory difficulties.

Materials. The N-Watch program (Davis, 2005) was used to select 24 low-frequency superset words (less than 30 occurrences per million) from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995), each of which contained a phonologically embedded subset equal or higher in frequency than the superset. The subset belonged to one of various semantic categories. The average difference in frequency between supersets and subsets was 53 occurrences per million.[1] Superset words comprised the subset plus one additional phoneme, which was positioned either before or after the subset. There were 12 items in each of the two embedding positions. Recordings were carried out in a sound-attenuated booth using a high-quality microphone and the CoolEdit Pro 2.0 software. Recordings were digitized (16 bit A/D) at 32 kHz. On output, they were converted to analog form (16 bit D/A, 32 kHz). Analyses of the waveforms of the distractors revealed that the embedded subsets (347 ms) were on average shorter than the isolated subsets (468 ms), which in turn were shorter than the supersets (519 ms). This is consistent with the findings of Salverda et al. (2003), and highlights the existence of subphonemic cues to the identity of the subsets. Each superset and corresponding subset was paired with a black-and-white line drawing, for example, the superset *couch* was paired with a picture of a *pig*, which is semantically associated with the subset cow. All of the 24 drawings were different. See Appendix A for a list of target pictures and the corresponding supersets and subsets along with their CELEX frequencies and the assignment of pictures and subsets to semantic categories.

Design. Each drawing was presented along with a spoken superset distractor in a semantically related condition (where its embedded subset belonged to the same semantic category as the drawing) and a semantically unrelated condition (where its subset belonged to a different category). In the unrelated condition, words and pictures from the same embedding condition (initial or final) were paired such that they were from different semantic categories. Each target-distractor pair was presented at two SOAs (−100 and 0 ms) measured from the onset of the embedded subset item. This is a range in which semantic effects are typically observed. The SOA conditions were blocked, and the order in which participants received the SOA conditions was counterbalanced. Within SOA, related and unrelated conditions were randomly intermixed. Each spoken superset word was presented once at each SOA in the related and unrelated conditions for a total of 96 trials (24 supersets × 2 SOAs × 2 relatedness conditions). In all, each picture was repeated seven times: four times in the critical experimental trials, and three times during the practice phase described below. It should be emphasized that all the picture-word pairs were semantically and phonologically unrelated (with half of the supersets in the critical trials including a semantically related subset; e.g., bus-scar). Furthermore, in practice blocks that included distractor words, the picture and distractors were unrelated. Thus, the relation between the pictures and critical words was quite opaque (much more so than most previous PWI experiments in which there is a transparent semantic or phonological relation between pictures and distractors). Accordingly, the present study should provide a measure of semantic activation of embedded words when strategic concerns are minimized.

Apparatus. The picture stimuli were presented as black line drawings over a white background on a computer monitor. Auditory distractors were played over good-quality headphones at approximately 65 dB SPL. Picture-naming responses were spoken into a microphone, and a voice key recorded response latencies to the nearest millisecond. Stimulus presentation and response recording were achieved using the DMDX software package (Forster & Forster, 2003).

Procedure. Each participant was tested individually. At the start of the experiment, participants were presented with a display of the entire picture set, along with the corresponding name printed below each picture, and instructed to familiarize themselves with each of the stimuli before commencing the experiment. Participants were then presented with two practice blocks, the first of which involved performing a naming response to each of the pictures without presentation of an auditory distractor, and the second involving performing naming responses to each of the pictures presented along with semantically unrelated auditory distractors. The practice trials were included to ensure that participants would retrieve the appropriate names in response to the pictures during the critical trials, which is a common procedure in PWI experiments. Immediately following the practice trials, the superset PWI condition was presented.

On each trial, a fixation cross appeared at the center of the screen for 500 ms. A blank screen then appeared for 500 ms, followed by the presentation of the picture and the distractor word. The auditory distractor was presented so that the onset of the embedded word was temporally aligned with the onset of the picture (0-ms SOA condition) or preceded the picture by 100 ms (−100-ms SOA condition). On each trial, the target picture remained onscreen for a maximum of 2 s, and the trial terminated either immediately after a naming response or if the participant failed to respond within 2 s. After each naming response, the experimenter judged the response to be either correct or incorrect. Incorrect responses included incorrect or incomplete responses, mouth clicks, and malfunctioning of equipment. Each trial was concluded by a 1-s intertrial interval.

Results

Trials in which participants gave an incorrect naming response, were disfluent, or produced a correction were excluded from the analysis. Similarly, trials in which RTs were less than 300 ms (these included trials where the voice key malfunctioned), or exceeded 1,500 ms were also removed from the analysis (in total 3.5% of trials were removed).

Table 1 shows the mean picture-naming latencies and errors across the various conditions. Because different sets of pictures and words were included in the initial and final embedding conditions, we analyzed these results separately. In all analyses, we employed linear mixed-effect modeling, which take into account the variability of both items and participants (Baayen, Davidson, & Bates, in press). Because the index of effect size η^2 cannot be straightforwardly applied to mixed-effect models, we used Cohen's *d* instead, which was obtained by dividing the estimated effect size of each effect or interaction by its pooled standard deviation.

---

[1] Supersets with equal or higher frequency subsets were included because, in previous work, we obtained evidence that written subset words are not activated to the level of meaning when they are lower frequency than the supersets (Bowers, Davis, & Hanley, 2005).

Table 1

Mean Picture Naming Latency (ms) and Error Rates (% in Parentheses) for Targets in Experiment 1 as a Function of the Semantic Relation of the Initial and Final Subsets to the Picture at SOAs of −100 ms and 0 ms

Embedding position	Distractor	SOA, ms	Semantic relatedness: Related	Semantic relatedness: Unrelated	Difference: RT	Difference: Err.
Initial (e.g., crayon)	e.g., PENce	−100	671 (2.8)	660 (2.4)	−11	−0.5
		0	688 (2.2)	655 (2.8)	-33^{***}	+0.7
Final (e.g., bus)	e.g., sCAR	−100	635 (5.7)	637 (1.7)	+2	-3.9^{***}
		0	650 (1.5)	636 (3.1)	−14	+1.6

Note. SOA = stimulus-onset asynchrony; RT = response time in milliseconds; Err. = error rate in percentage.
*** $p < .001$.

For the initial embedding condition, the analysis carried out on response latencies, with subjects and items as random factors and semantic relatedness and SOA as fixed factors, showed an effect of relatedness $F(1, 1773) = 10.15$, $p = .001$, $d = .13$, reflecting longer response times in the related than unrelated conditions. SOA was not significant, $F(1, 1773) = 1.00$, $p = .32$, nor was the interaction between relatedness and SOA, $F(1, 1773) = 1.83$, $p = .18$. Parallel analyses carried out on the error data failed to show any effects, with all Fs < 1. Accordingly, the relatedness effect observed in the RTs was not compromised by a speed-accuracy trade-off.

For the final embedding condition, the analysis of RTs failed to show a main effect of relatedness, $F(1, 1765) < 1$. Furthermore, neither the main effect of SOA nor the interaction between SOA and relatedness approached significance, $F(1, 1765) = 1.36$, $p = .24$, and $F(1, 1765) = 1.28$, $p = .26$, respectively. The analysis of the error data also failed to show a main effect of relatedness, $F(1, 1820) = 2.42$, $p = .12$, whereas the effect of SOA approached significance, $F(1, 1820) = 3.38$, $p = .07$, $d = .08$. However, the interaction between relatedness and SOA was highly significant, $F(1, 1820) = 12.48$, $p < .001$, $d = .34$, which reflects the strong semantic interference at SOA −100 ms.

To explore the PWI effects more directly, analyses were carried out separately at each SOA. As indicated in Table 1, a highly significant interference effect was obtained in the RTs for initial embedded words at SOA 0, and a corresponding effect was obtained in the errors for final-embedded words at SOA −100.

Discussion

The main finding from Experiment 1 is that picture naming is impaired by a semantically related spoken word embedded within an unrelated superset. This PWI effect was obtained in the RT analysis when subsets were in initial (*PENce*) position and in the error analysis when subsets were in final (*sCAR*) position. The latter finding challenges some previous studies that have failed to obtain evidence for the coactivation of final embedded subsets (e.g., Norris et al., 2006). Furthermore, in the studies that have provided evidence for the coactivation of final embedded subsets, the effects have been restricted to conditions in which the subsets were marked sublexically by a syllable boundary (e.g., Vroomen & de Gelder, 1997). In the current study, however, although the final subsets were unmarked, we nevertheless obtained a PWI effect. Thus, the present study extends the set of conditions in which embedded words are activated.

Why do we obtain evidence for more widespread activation of final embedded words compared to previous research? Apart from the difference in tasks, it is worth noting that the PWI effect for final-embedded words was only obtained in the −100-ms SOA condition (i.e., the picture was presented 100 ms after the onset of the embedded words, such that the picture and the spoken superset overlapped). The restriction of PWI effects to one SOA condition raises the possibility that the activation time course of subsets is highly constrained, and perhaps difficult to detect. In this regard, it is worth noting that the Norris et al. (2006) study presented the visual word primes either at the offset of the spoken word or with a 500-ms interstimulus interval. Similarly, Gow and Gordon (1995), Marslen-Wilson et al. (1994), and Vroomen and de Gelder (1997) presented the written words at the offset of the spoken target. This suggests that more priming could have been obtained in these studies if the presentation of the spoken and written words had overlapped. Another difference is that our subsets were all highly imageable words, and given that lexical coactivation is often measured through semantic tasks (typically cross-modal semantic priming), it is possible that imageable words provide a more sensitive measure of lexical coactivation than less imageable ones. Therefore, again, it is likely that PWI effects for final embedded words would be more robust in priming paradigms if the subsets were restricted to highly imageable words.

Whatever the reason for the discrepancy, the significant PWI effect for final embedded items at SOA −100 ms poses a challenge to the currently implemented Shortlist model, which predicts little or no activation of final embedded words. By contrast, our results are consistent with the TRACE model that does predict some activation of these items. TRACE also predicts larger PWI effects for initial than final subsets. However, it is difficult to make comparisons between the initial and final effects reported here, because different words and pictures were used across conditions and because the PWI effects were obtained in the RT analyses in one condition and in the error analyses in the other.

Experiment 2

The stimuli of Experiment 1 were selected such that subsets and supersets differed by a single phoneme. This may have contributed to the PWI effects observed for the word-final subsets because, under such conditions, there is little time to activate the superset before the encoding of the subset itself. For example, when participants heard the word *scar*, the superset *scar* could have been only minimally activated before the beginning of the subset *car*. At this early stage of processing, the superset may not be an effective competitor of the subset. By contrast, when supersets and subsets differ by several phonemes, it is possible that the supersets are more powerful competitors. For example, when presented with the word *membrane*, the superset *membrane* may be active and an effective competitor to *rain* or *brain* before the subset receives any bottom-up activation. In addition, given that subsets and supersets in Experiment 1 differed by a single phoneme, our evidence for embedded activation could be reduced to a case of neighborhood activation (Luce & Pisoni, 1998). Indeed, in TRACE simulations, subsets are more activated when subsets and supersets differ in few compared to many phonemes (Frauenfelder & Peeters, 1990).

In Experiment 2, we assessed whether final subsets are activated when their supersets include three or more additional phonemes. Under these conditions, there is more time to activate the superset before the onset of the subset, and the relative degree of phonological overlap between the superset and the subset is reduced. In addition to manipulating the number of mismatching phonemes, we varied the syllabic status of the embedded words. We included final subsets that were themselves syllables within the supersets (e.g., *motoRIST*, *wrist*) and final subsets whose onsets did not align with a syllable boundary (e.g., *squATTER*, *otter*).

Method

Participants. Twenty-six undergraduate students from the University of Bristol, none of whom had participated in the first experiment, were recruited and paid a small honorarium for their participation. All were native speakers of British English, had normal or corrected-to-normal eyesight, and no self-reported auditory difficulties.

Materials. Twenty-five low-frequency superset words (CELEX frequency of less than 15 occurrences per million) were selected as auditory distractors. As in Experiment 1, these words contained an embedded subset equal or higher in frequency relative to the superset word. The mean difference in frequency between subsets and supersets was 32 occurrences per million. The subset belonged to one of several semantic categories. Superset words comprised the subset plus three to four additional phonemes, positioned before the embedded subset. Additionally, supersets were subclassified into two groups: aligned-syllable subset stimuli, in which the onset of the embedded word was aligned with a syllable onset in the superset (e.g., selFISH); and misaligned-syllable subset stimuli, in which the onset of the subset did not coincide with the onset of a syllable in the superset (e.g., resTRAIN). As in Experiment 1, the embedded subsets were on average shorter in duration than the isolated subsets, with the supersets the longest. This was the case for the aligned stimuli (370, 466, and 712 ms, respectively), as well as for the misaligned stimuli (380, 477, and 701 ms, respectively). Each superset was paired with a black-and-white line drawing belonging to the same semantic category as its subset. All of the 25 drawings were different. See Appendix B for a list of target pictures and the corresponding supersets and subsets along with their CELEX frequency and the assignment of pictures and subsets to semantic categories.

Design and procedure. These were the same as in Experiment 1.

Results

Trials in which participants gave an incorrect naming response, were disfluent, or produced a correction were excluded from the analysis. Similarly, trials in which RTs were less than 300 ms (these included trials where the voice key malfunctioned) or exceeded 1,500 ms were considered outliers and were also removed from the analysis (5.2% of all trials). The data from one participant were excluded because of a 44% error rate.

Table 2 shows the mean picture naming latencies and errors across the various conditions. For the aligned-syllable embedded condition, an analysis on response latencies, with subjects and items as random factors and semantic relatedness and SOA as fixed factors, showed an effect of relatedness $F(1, 1166) = 4.61$, $p < .05$, $d = .09$, reflecting slower RTs in the related than unrelated conditions. SOA was also significant, $F(1, 1166) = 4.19$, $p < .05$, $d = .11$, with shorter naming latencies at the −100-ms than 0-ms SOA conditions. Finally, there was no interaction between relatedness and SOA, $F(1, 1166) = 1.41$, $p = .23$. In the analysis by errors, there was no effect of relatedness and no interaction between relatedness and SOA, with F values < 1. The effect of SOA only approached significance, $F(1, 1244) = 2.09$, $p = .15$.

For the misaligned-syllable embedded condition, an analysis of the RTs revealed no effect of semantic relatedness, $F(1, 1282) <$ 1, and no interaction between relatedness and SOA, $F(1, 1282) <$ 1. However, the effect of SOA was significant, $F(1, 1282) = 8.59$, $p < .01$, $d = .13$, reflecting the faster RTs in the −100-ms than the 0-ms SOA conditions. The analysis of errors showed no significant

Table 2

Mean Picture Naming Latency (ms) and Error Rates (% in Parentheses) for Targets in Experiment 2 as a Function of Syllable Status of the Final Subset and Semantic Relation of the Subset to the Picture at SOAs of −100 and 0 ms

			Semantic relatedness		Difference	
Embedding position	Distractor	SOA, ms	Related	Unrelated	RT	Err.
Aligned (e.g., bowl)	e.g., chilDISH	−100	705 (5.1)	680 (5.4)	-25^{*}	+0.3
		0	713 (6.7)	706 (7.7)	−7	+1.0
Misaligned (e.g., bus)	e.g., resTRAIN	−100	653 (5.3)	650 (4.4)	−3	−0.9
		0	673 (5.6)	667 (4.1)	−6	−1.5

Note. SOA = stimulus-onset asynchrony; RT = response time in milliseconds; Err. = error rate in percentage.
$^{*}p < .05$.

effect of relatedness, $F(1, 1348) = 1.11$, $p = .29$, SOA, $F(1, 1348) < 1$, or interaction, $F(1, 1348) < 1$.

To explore the PWI effects in more detail, analyses were carried out at each SOA condition. As can be seen in Table 2, a PWI effect was obtained only in the aligned-syllable condition at −100 SOA.

Discussion

The main result from Experiment 2 was that a PWI effect was obtained for final embedded words that differed from their supersets by three or more phonemes. However, in contrast with Experiment 1, the effect was restricted to conditions in which the onset of the embedded word was aligned with the onset of a syllable in the superset. In this respect, it appears that long supersets are more effective in suppressing the activation of final embedded subsets than shorter ones are.

Experiment 3

In Experiments 1 and 2, we obtained clear evidence for embedded-word activation with the PWI procedure, but it is necessary to compare the amount of PWI produced by the embedded words with that of the same words presented in isolation. This will provide an estimate of the relative degree of activation for embedding and embedded words. In the extreme case, PWI effects for embedded words could match the effects obtained for the isolated words, which would indicate that the embedded words are fully activated. In Experiment 3, we assessed PWI effects for the isolated subsets of Experiments 1 and 2.

Method

Participants. Eighteen participants, none of whom had been in the first two experiments, were recruited and paid a small honorarium for their participation. All were native speakers of British English, had normal or corrected-to-normal eyesight, and no self-reported auditory difficulties.

Materials. The materials consisted of the subset words from Experiments 1 and 2, paired with the same semantically related black-and-white line drawings used in Experiments 1 and 2.

Design and procedure. Line drawings were presented along with auditory subset distractors in semantically related and unrelated conditions. All other aspects of the design and procedure were the same as before.

Results and Discussion

Trials in which participants gave an incorrect naming response, were disfluent, or produced a correction, were excluded from the analysis (1.0% of trials). Trials in which RTs were less than 300 ms (including trials where the voice key malfunctioned) or exceeded 1,500 ms were removed from the analysis (6.2% of trials, for a total error rate of 7.2%).

Table 3 shows the mean picture naming latencies and errors for the subsets across the various conditions. We first ran a linear mixed-effect analysis on response latencies from the subsets taken from Experiment 1, with participants and items as random factors and semantic relatedness and SOA as fixed factors. An analysis on the initial embedded items showed an effect of relatedness in the RTs, $F(1, 800) = 6.71$, $p < .01$, $d = .14$, reflecting slower RTs in the related than the unrelated conditions. There was no SOA effect, $F(1, 800) = 1.35$, $p = .24$, or interaction, $F(1, 800) < 1$. An analysis of the final embedded words failed to show an effect of relatedness, $F(1, 806) < 1$ or SOA, $F(1, 806) = 1.66$, $p = .20$, but the interaction between relatedness and SOA approached significance, $F(1, 806) = 3.61$, $p = .06$, $d = .24$, which was the product of a PWI effect at SOA 0 and a small tendency for a facilitatory effect at SOA −100. The basis of the facilitatory effect is unclear. None of the accuracy analyses approached significance. To look at the PWI effects more closely we calculated the effects at each SOA condition. As can be seen in Table 3, a PWI effect was obtained for both initial and final embedded words.

A similar analysis on the subsets taken from Experiment 2 revealed an effect of relatedness on the RTs of the aligned condition, $F(1,792) = 23.47$, $p < .001$, $d = .22$. None of the other effects approached significance, except for the interaction between relatedness and SOA in the error rates, $F(1, 860) = 3.21$, $p = .07$, $d = .26$, reflecting a trend toward a PWI effect at SOA 0. An analysis of the misaligned subsets revealed a main effect of relatedness, $F(1, 861) = 11.72$, $p < .001$, $d = .16$, with no other effects approaching significance. Relatedness effects were also calculated at each SOA condition. As can be seen in Table 3, a PWI effect was obtained in all RT conditions.

Table 3

Mean Picture Naming Latency (ms) and Error Rates (% in Parentheses) for Targets in Experiment 3 as a Function of Semantic Relation of the Subset to the Picture at SOAs of −100 and 0 ms

			Semantic relatedness		Difference	
Replication	Condition	SOA, ms	Related	Unrelated	RT, ms	Err., %
Experiment 1	Initial	−100	766 (4.6)	739 (6.5)	-27^{*}	+1.9
		0	769 (8.3)	750 (8.3)	$-19^{\dagger}$	0.0
	Final	−100	746 (5.1)	759 (4.2)	+13	−0.9
		0	775 (7.9)	746 (7.9)	-29^{*}	0.0
Experiment 2	Aligned	−100	800 (6.9)	757 (8.8)	-43^{***}	+1.9
		0	797 (10.2)	756 (5.6)	-41^{**}	$-4.6^{\dagger}$
	Misaligned	−100	756 (9.8)	734 (7.3)	-22^{*}	−2.5
		0	760 (6.4)	721 (6.8)	-39^{**}	+0.4

Note. SOA = stimulus-onset asynchrony; RT = response time in milliseconds; Err. = error rate in percentage.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

The critical issue, however, is the size of the PWI effect for the subsets presented in isolation (Experiment 3) relative to that for the corresponding supersets (Experiments 1 and 2). In the case of the subsets and supersets that differed by a single phoneme (Experiment 1), similar PWI effects were obtained: Collapsing across SOA and position, the overall PWI effect was 16 ms for the supersets in Experiment 3 and 14 ms for the subsets in Experiment 1, $F(1, 5156) < 1$. Similarly, there was no difference in errors, $F(1, 5372) < 1$. This suggests that subset words (e.g., *car*) were activated to a similar degree when they were embedded in superset words (e.g., *sCAR*) or presented in isolation. Thus, supersets do not appear to inhibit the activation of their subsets when the supersets and subsets differ by a single phoneme, at least not in the early stages of word processing.

In the case of the isolated subsets and corresponding multisyllable supersets (Experiment 2), larger PWI effects were obtained for the subsets. In the aligned-syllable condition, the overall PWI effect was 43 ms in Experiment 3 and 16 ms in Experiment 2. A linear mixed-effect analysis run on response latencies, with subjects and items as random factors, and Relatedness and Experiment as fixed factors, showed an effect of Relatedness, $F(1, 1962) = 22.17$, $p < .001$, $d = .14$, an effect of Experiment, $F(1, 1962) = 5.74$, $p < .05$, $d = 41$, and critically, a significant interaction, $F(1, 1962) = 5.68$, $p < .05$, $d = .20$. That is, when supersets and subsets differed by several phonemes, the supersets acted to reduce the activation of the embedded words even when they constituted a syllable of the superset. In the case of the subsets in the misaligned condition, the overall PWI effect was 30 ms in Experiment 3 and was lost in Experiment 2 (5 ms): Relatedness effect, $F(1, 2147) = 9.21$, $p < .005$, $d = .09$; Experiment effect, $F(1, 2147) = 5.89$, $p < .05$, $d = 48$; interaction, $F(1, 2147) = 6.81$, $p < .01$, $d = 15$.

Although the results of Experiment 2 suggest that final embedded subsets within multiple syllable supersets are only activated when they are syllable aligned, it should be noted that the PWI effect obtained in Experiment 3 for subsets in the misaligned condition (30 ms) was somewhat smaller than in the aligned condition (43 ms). That is, the reduced PWI effects in the nonaligned condition in Experiment 2 might not simply reflect the syllable structure of the supersets, but also the reduced effectiveness of the subsets per se. Although an analysis of Experiment 3 that included Subset Type (aligned vs. misaligned) and Relatedness did not show a significant interaction, $F(1, 1657) = 1.31$, $p = .25$, it is still possible that if the subsets in the misaligned condition were better at producing PWI effects, a larger PWI effect would have been found for the misaligned condition in Experiment 2.

To provide a stronger test of the claim that misaligned subsets do not support a PWI effect, we rank ordered misaligned subset-superset pairs from Experiment 3 in terms of the size of the PWI effect they produced at SOA -100 (the condition that supported a PWI effect in Experiment 2). After removing three pairs (*eye-adhere*, *horse-diagram*, and *sun-constrain*) that showed the least interference, the PWI effects in Experiment 3 were 41 ms at SOA -100 and 48-ms effect at SOA 0 (similar to the corresponding PWI effects for the aligned condition). We then examined the effect of removing the same three pairs from the misaligned condition of Experiment 2. There was relatively little effect: In the SOA -100 condition, the PWI effect changed from 3 ms to 10 ms ($p = .50$), while in the SOA 0 condition the PWI effect changed from 6 ms to 7 ms ($p = .40$). As before, the overall PWI effect was not significant, $F(1, 995) = 1.21$, $p = .27$. Thus, the absence of a PWI effect for the misaligned subsets of Experiment 2 (e.g., resTRAIN when naming a bus) continued to be observed even when the same words presented in isolation (in Experiment 3, e.g., TRAIN when naming a bus) showed PWI effects that were equivalent to those for the embedded words from the aligned subset condition.

General Discussion

The current study provides evidence in support of the claim that initial and final embedded subset words are activated during the process of identifying spoken supersets (e.g., *bone* in *trombone*). The critical finding in Experiment 1 is that final-embedded subsets are activated in the absence of any syllable-boundary cues when the subsets and supersets differ by a single phoneme (e.g., *car* in *scar*). This contradicts the claim that final embedded subsets are only activated when their onset is marked by a syllable boundary (Gow & Gordon, 1995; Vroomen & de Gelder, 1997). In Experiment 2, the subsets and supersets differed by three or more phonemes and, again, final embedded words were activated. However, in this case, activation was restricted to conditions in which the final subsets were aligned with a syllable boundary (*bat* in *acrobat*); little or no activation was obtained for subsets not aligned with a syllable boundary (*ear* in *pioneer*).

Apart from highlighting the extent to which embedded words are activated, the results provide the first demonstration that the amount of overlap between subsets and supersets plays a key role in constraining the activation of final embedded items (consistent with the prediction of TRACE; Frauenfelder & Peeters, 1990). Evidence for this comes from two sources. First, as noted above, final embedded subsets misaligned with the syllable structure of supersets were only activated when the subsets and supersets differed by a single phoneme (of course, the final embedded subsets for monosyllable supersets are necessarily misaligned). Second, similar levels of PWI were obtained for the supersets in Experiment 1 and the corresponding isolated subsets in Experiment 3, reflecting a strong degree of activation for those stimuli, whereas the supersets in the aligned-syllable condition in Experiment 2 produced approximately one third of the interference of the corresponding isolated subsets in Experiment 3.

These findings suggest that a subset word like *ear* is more strongly activated in a short word like *near* than in a longer word like *pioneer*. This effect of superset word length is consistent with models of speech perception that incorporate lexical competition (e.g., McClelland & Elman, 1986; Norris, 1994). In such models, there are at least two factors that may cause units that code long words to attain larger activities than those that code short words. First, long words are supported by more bottom-up evidence than shorter words (e.g., *pioneer* is supported by six phonemes, compared to only three phonemes for *near*). Second, short words may experience greater inhibition than long words because they are similar to more words. For example, the set of phonological neighbors for *near* includes *beer*, *dear*, *fear*, *gear*, *hear*, *jeer*, *leer*, *mere*, *tier*, *weir*, *nor*, etc. (the N-Watch program lists 19 phonological neighbors for *near*), whereas *pioneer* has no phonological neighbors except for its inflection *pioneers*. It has also been suggested that competitive network models should be designed such that lexical units that code longer words output stronger inhibitory signals than units that code shorter words (Davis, 1999; Grossberg, 1986) to ensure that the supersets rather than subsets are most strongly activated in response to a superset input. It is relevant to note that an advantage for long words over shorter words has recently been reported in another paradigm for studying speech perception. Pitt and Samuel (2006) found that lexical effects on phoneme perception (as measured by the lexical shift associated with the Ganong effect) were substantially stronger and more robust for long words than for short words. They interpreted this as evidence in favor of competitive network models in which lexical activity is greater in long than short words, develops earlier, and remains longer. This pattern would also imply that longer words are better able to suppress their subset competitors, as observed in the present experiments.

Perhaps the most surprising finding is that the supersets in Experiment 1 and the corresponding subsets in Experiment 3 produced equal PWI effects. This would seem to suggest that these final embedded subsets were activated fully, without any inhibition from the supersets. This happened despite the fact that the onset of the embedded words was not marked by a syllable boundary, and despite the fact that the embedded subsets were acoustically shorter in duration than the corresponding subsets presented in isolation. As noted above, acoustic duration is a subphonemic cue that has been found to modulate the extent to which embedded words are activated (Salverda et al., 2003). Although we take our findings to provide strong evidence that initial and final embedded words are activated when the subsets and supersets differ by a single phoneme, we are hesitant to conclude that there is no inhibition from the supersets to the subsets. Indeed, the subsets must at some point be suppressed, as listeners usually identify supersets and not their subsets. However, the findings do highlight the weak nature of subset/superset competition in the early stages of processing under these conditions.

Comparison Between the PWI Task and Previous Methods for Measuring Coactivation of Form-Related Words

In addition to providing evidence regarding the conditions in which embedded words are activated, the current work introduces a new method for studying spoken-word recognition. One obvious question is whether this task has any advantages over previous ones. We suggest that it does.

Cross-modal associative priming. As noted in the Introduction, the most common method for assessing subset/superset coactivation has been the cross-modal associative priming paradigm, in which participants hear a spoken word (e.g., *trombone*) and make a lexical decision to a written word that is semantically related to the embedded subset (e.g., *rib*). An advantage of this approach is that experiments can be designed so that participants respond to the same visual target in the related and unrelated prime conditions (e.g., participants respond to the written word *rib* when presented with the related and unrelated spoken words); and when designed this way, any priming effects cannot be attributed to any uncontrolled differences in the target stimuli. This is a methodological improvement over studies that compared the identification of words with and without subsets, a procedure that requires that different words be used across conditions (Cluff & Luce, 1990; Luce & Lyons, 1999). In the latter studies, the authors attempted to match words in all relevant respects other than embeddedness, but this proved a difficult (if not impossible) task (cf., Bowers, Davis, & Hanley, 2005; Lewis, 2006).

Although the cross-modal priming procedure has provided important (albeit somewhat inconsistent) findings concerning the activation of embedded words, we suggest that the PWI procedure enjoys some advantages. Most notably, picture naming is a relatively natural task, and unlike the lexical-decision task used in the cross-modal priming studies, it does not involve meta-linguistic decisions (e.g., categorizing a written letter string as a word or a nonword). Furthermore, the interference obtained in the PWI task is typically considered a form of Stroop interference (e.g., Roelofs, 2003), and thus may be less subject to strategic effects than paradigms involving facilitatory priming results. It also shares the advantage of cross-modal priming that participants respond to the same target in both related and unrelated conditions.

Visual-world paradigm. Recently, there has been an explosion of research employing the visual-world paradigm to address a wide range of psycholinguistic issues (e.g., Tanenhaus & Brown-Schmidt, in press). In this procedure, a small number of objects (typically four; the so-called *visual world*) are displayed on a computer monitor, and participants are instructed to point to or look at one of the objects. Eye movements are monitored, and the proportion of looks to each of the objects is calculated at brief intervals (e.g., every 33 ms) following the onset of the critical spoken word. For instance, a participant might be asked to point to a picture of a *speaker*, and eye-movements to the target picture (*speaker*), related foils (e.g., a picture of a *beaker*, which rhymes with *speaker*), and unrelated foils are monitored at each time slice. An increased proportion of fixations to a related compared to an unrelated foil is taken to indicate that the spoken word has activated the lexical-phonological and semantic representations of the foil. One key advantage of this approach is that the analysis of the fixations across the time slices provides a measure of the time course of the activation process.[2] Another advantage is that participants are simply asked to look (and sometimes point) to objects that are named; that is, the task is often considered ecologically valid.

Most relevant for present concerns, the visual-world paradigm provides evidence that initial embedded words are activated to the level of meaning (Salverda et al., 2003), and a number of studies have provided evidence that cohort- and rhyme-related words are also activated semantically. For instance, when the target word *speaker* is presented, there is an increased proportion of eye-fixations to a picture of a *beaker* compared to that of an unrelated object (Allopenna, Magnuson, & Tanenhaus, 1998). We consider this task in some detail given the recent interest in the approach.

---

[2] The ability to analyze responses across a series of time slices is often described as an advantage of the visual world paradigm compared to other methods. However, it should be noted that there is nothing special about the procedure in this regards. Eye movements are ballistic (just as key presses or spoken utterances) and an analysis of responses across a range of RTs can theoretically be performed in all tasks. For example, Balota and colleagues have used vincentile analyses to great effect to study the relative time course that frequency, visual degradation, semantic variables, and so forth, have on reading written words in the lexical-decision and naming tasks (e.g., Yap & Balota, 2007).

One of the claimed advantages of the visual-world compared to the cross-modal semantic priming paradigm is that it provides a more sensitive measure of lexical coactivation. This may explain why the visual-world paradigm shows that rhyme-related words are activated to the level of meaning, whereas priming studies provide little or no evidence for this claim (e.g., Connine, Blasko, & Titone, 1993; Marslen-Wilson & Zwitserlood, 1989). Indeed, Allopenna et al. (1998) argue that: ". . . semantic priming may be too insensitive or too indirect a response measure to reveal activation of rhymes." Another possibility, however, is that the visual-world paradigm distorts the normal process of identifying spoken words (e.g., by preactivating the lexical phonological forms of words via priming from the pictures), or induces strategies for interpreting the words within the experimental context (e.g., participants may be more likely to interpret the spoken target *speaker* as *beaker* when it is preceded by the picture of a *beaker*). In both of these scenarios, the procedure would produce misleading evidence about the extent to which words are coactivated during speech processing in naturalistic contexts.

Advocates of the visual-world paradigm are of course aware of these concerns, and a number of findings have been presented as evidence that the process of activating the lexical-phonological forms of words is unaffected by the array of pictures that precedes the spoken target. For example, Allopenna et al. (1998) found that the TRACE model of word identification and their own behavioral results produce similar estimates of the pattern of coactivation across cohorts and rhymes (see also Dahan, Magnuson, Tanenhaus, & Hogan, 2001). Given that lexical activation in the model was not biased by any pictorial or semantic context, human performance might be interpreted in the same way. In addition, a number of studies have shown that eye-tracking is impacted by the phonological properties of words that are *not* depicted in the visual world. For instance, participants are slower to fixate a target picture when the target is from a high- than low-density neighborhood, and this occurs even when the distractor pictures are all unrelated to the target (e.g., Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Magnuson, Dixon, Tanenhaus, & Aslin, 2007). According to Tanenhaus and colleagues (e.g., Dahan & Tanenhaus, 2004; Tanenhaus & Brown-Schmidt, in press), this provides the strongest evidence that the visual-word context does not distort the normal course of identifying spoken words.

Still, there are reasons to be cautious in adopting this conclusion. First, the fact that TRACE shows rhyme-related activation in the absence of a constrained visual display should not be taken as evidence that the behavioral results in the visual-world paradigm are similarly unconstrained by the visual display. The normal chain of logic is to use an experimental result as evidence for a theory, but here a theory is used to support the results obtained with a particular method. This line of reasoning is circular, and the logic can be used to support quite different conclusions. For example, given that Shortlist predicts that final embedded words are not coactivated, reports of null priming for final embedded words (e.g., Norris et al., 2006) could be used to argue that the cross-modal semantic priming procedure provides an accurate measure of lexical coactivation. Second, although we agree that the neighborhood effect reported by Magnuson et al. (2003, 2007) shows that the word set activated by the target is not restricted to the displayed pictures, it does not provide evidence that picture contexts are irrelevant to the processing of the target in general. It is quite possible (and seems plausible to us) that the target *speaker* is more likely to be perceived (or interpreted) as *beaker* in the context of the picture of a *beaker*, and at the same time, that various perceptual and lexical factors (e.g., neighborhood density) could affect the identification of *speaker* in the absence of a *beaker*. That is, bottom-up constraints such as neighborhood density, which operate independently of the visual world, and top-down constraints such as context, which is imposed by the visual world, can both influence word identification.

Most importantly, however, the claim that the restrictiveness of the visual display does not misrepresent genuine speech processing seems at odds with some related findings. For example, sentential contexts do influence the on-line activation of spoken words in the visual-world paradigm. Dahan and Tanenhaus (2004) assessed eye fixation to targets and cohort-related pictures when a spoken target was presented either in a neutral context or in a verb context that favored the target. Participants fixated the cohort competitors more often than a control item in the neutral-context condition only. Based on this finding, the authors claimed that sentential context can eliminate the activation of cohort competitors. That is, the sentence context radically altered which word forms were activated in response to a spoken word. Yet, as noted above, to make claims about the coactivation of word forms in natural contexts, it needs to be claimed that the visual world restricted to one of four alternatives does *not* influence speech processing. Why one prior context (a sentential context) should eliminate cohort activation while another prior context (pictures) should have no impact on coactivation is unclear.

An alternative defense of the visual-world paradigm is to concede that the previewed objects do influence lexical activation, but to argue that objects in the real world do the same. We find this possibility unlikely: In the real world, listeners rarely know in advance that an upcoming spoken word is guaranteed to be one of a very small number of possibilities. Thus, the top-down influences generated by a visual-world display are unlikely to produce effects that are characteristic of normal word processing.

We do not mean to suggest that the visual-world paradigm is poorly suited to address all questions. In our view, the problem arises when inferences about spoken-word recognition are based on (mis)fixations to pictures that share phonological, visual, or semantic features with the target. When inferences depend upon eye fixations to target pictures surrounded by unrelated foils, there is less reason to assume that the visual world has an artificially restrictive effect. A number of important findings have been reported in exactly these contexts (e.g., Dahan et al., 2001; Magnuson et al., 2007).

Returning to our comparison with PWI, a key advantage of the PWI procedure is that it seems less likely to bias perception or induce strategic effects than other paradigms. In the PWI task, the spoken words and the pictures are phonologically unrelated to one another (e.g., naming the picture of a *bus* while hearing the word *scar*) and, critically, the spoken words either are presented at the same time as the pictures or they *precede* the pictures. This constitutes a clear advantage over the visual-world paradigm, in which the pictures (e.g., *beaker*) and the spoken words (*speaker*) are phonologically related, and the pictures are presented long before the spoken words (3 s in Allopenna et al., 1998). Although the time interval between the onset of the visual world and the onset of the spoken target is often shorter (e.g., less than 500 ms in Magnuson et al., 2007), a relatively long time interval between the pictures and words is required so that participants are aware of the identity and location of the pictures, which in turn allows eye movements to be informative about language processes rather than being the product of blind visual exploration.

What Is the Basis of the Interference in the PWI Task?

Before concluding, it is worth considering the nature of the representations and processes supporting the current PWI effects. It is tempting to conclude that the subset words (e.g., *bone* in *trombone*) were activated to the level of meaning and interfered with the naming of the related pictures (*skull*). Indeed, it is commonly argued that semantic interference in the PWI paradigm is because of competition between the lexical-semantic representations activated by the pictures and the spoken-word forms (e.g., Damian & Bowers, 2003). The cross-modal associative priming effects reviewed above (e.g., the spoken word *trombone* priming the written word *rib,* Shillcock, 1990) also supports the claim that spoken subsets can be activated to the level of meaning.

If this is the case, it implies that speech processing is cascaded across the phonological-semantic interface. That is, the lexical-phonology of a subset word is activated, and this in turn produces semantic activation of the subset before its phonology being suppressed by the superset word. However, based on findings from the speech production literature, it is possible to construct an argument in which the current PWI effects are the product of the superset distractors activating the lexical–phonological representations of the subsets rather than their semantic representations. It is often claimed that speech production is cascaded (e.g., Dell, 1986; McQueen, Dahan, & Cutler, 2003), with the conceptual representation of a to-be-named object (e.g., *cat*) activating a cohort of lexical-semantic representations, and in turn activating a cohort of phonological forms, with phonological activation a function of the degree to which an item is activated semantically (e.g., the phonological form of *cat* would be more active than *dog*). This hypothesis leaves open the possibility that the interference we obtained reflects cascaded processing in speech production rather than speech perception. For example, the picture naming of *skull* could be impaired by the spoken distractor *trombone* because of the lexical-phonology of *bone* being highly activated because of both bottom-up input (from *trombone*) and top-down cascaded processing (from *skull*). This, in principle, could lead to increased competition at the lexical-phonological level, making it more difficult to select *skull* even though the semantics of bone was never activated from *trombone*. Indeed, this is the sort of argument that has been advanced to argue for cascaded processing in speech production (e.g., Jescheniak, Hahne, Hoffmann, & Wagner, 2006; Jescheniak & Schriefers, 1998).

Given these claims for speech production, the current PWI results do not rule out a modular (noncascaded) account of speech perception in which semantic activation is restricted to the identified word (presumably the superset). However, the findings do provide strong evidence that either speech production *or* speech perception (or both) is cascaded across the semantic-phonological interface. We suggest that it is more parsimonious to assume that both perception and production are cascaded than to assume qualitative differences between these linked processes.

Whatever the conclusion one draws regarding cascaded processing from the PWI results, the current findings provide direct evidence that the lexical-phonology of subset words (e.g., *bone*) is activated from spoken supersets (e.g., *trombone*), contrary to the claim of Norris et al. (2006). Thus, even if speech processing is modular, and there is no semantic activation of *bone* from the spoken word *trombone*, the current PWI effects nevertheless depend on the lexical-phonology of *bone* being activated based on the input *trombone*. Critically, this activation is unlikely to be the by-product of strategies induced by the pictures, given that the pictures and spoken words are presented simultaneously (or the words preceded the pictures). Therefore, we argue that the current study provides the strongest evidence to date that initial and final subsets are activated to the level of lexical-phonology during the course of identifying spoken words.

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.

Baayen, R. H., Davidson, D. J., & Bates, D. M (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Bowers, J. S., Davis, C. J., & Hanley, D. (2005). Is there a “hat” in “that”? *Journal of Memory and Language, 52,* 131–143.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 551–563.

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

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133–142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16,* 507–534.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Evidence from immediate effects of verb-based constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30,* 498–513.

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin and Review, 10,* 111–117.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1–18.

Davis, C. J. (1999). The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition (Doctoral dissertation, University of New South Wales). *Dissertation Abstracts International, 62,* 594. (http://www.pc.rhul.ac.uk/staff/c.davis/Thesis/thesis.zip).

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods, 37,* 65–70.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 218–244.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Forster, K. I., & Forster, J. C. (2003). A Windows display program with millisecond accuracy. *Behavior Research Methods Instruments and Computers, 35,* 116–124.

Frauenfelder, U. H., & Peeters, G. (1990). On lexical segmentation in TRACE: An exercise in simulation. In G. Altmann (Eds.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 50–86). Cambridge, MA: MIT Prss.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 344–359.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol. 1. Speech perception.* New York: Academic Press.

Isel, F., & Bacri, N. (1999). Spoken-word recognition: The access to embedded words. *Brain and Language, 68,* 61–67.

Jescheniak, J. D., Hahne, A., Hoffmann, S., & Wagner, V. (2006). Phonological activation of category coordinates during speech planning is observable in children but not in adults: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 373–386.

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory and Cognition, 5,* 1256–1274.

Lewis, M. B. (2006). Chasing psycholinguistic effects: A cautionary tale. *Visual Cognition, 13,* 1012–1026.

Luce, P. A., & Cluff, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception and Psychophysics, 60,* 484–490.

Luce, P. A., & Lyons, E. A. (1999). Processing lexically embedded spoken words. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 174–183.

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

Lupker, S. J., & Sanders, M. (1982). Visual-field differences in picture-word interference. *Brain and Cognition, 1,* 381–398.

Magnuson, J. S., Dixon, J., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31,* 133–156.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word recognition and learning: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132,* 202–227.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101,* 3–33.

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

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134,* 477–500.

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

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes, 10,* 309–331.

McQueen, J. M., Dahan, D., & Cutler, A. (2003). Continuity and gradedness in speech processing. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production. Differences and similarities* (pp. 39–78). Berlin: Mouton de Gruyter.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20,* 298–305.

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

Norris, D. G., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology, 53,* 146–193.

Norris, D. G., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34,* 191–243.

Pitt, M. A., & Samuel, A. G. (2006). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1120–1135.

Posnansky, C. J., & Rayner, K. (1977). Visual-feature and response components in a picture-word interference task with beginning and skilled readers. *Journal of Experimental Child Psychology, 24,* 440–460.

Prather, P., & Swinney, D. (1977). Some effects of syntactic context upon lexical access. Paper presented at the *85th Annual Convention of the American Psychological Association.* San Francisco, CA.

Rayner, K., & Springer, C. J. (1986). Graphemic and semantic similarity effects in the picture-word interference task. *British Journal of Psychology, 77,* 207–222.

Rodd, J. (2004). When do leotards get their spots? Semantic activation of lexical neighbours in visual word recognition. *Psychonomic Bulletin and Review, 11,* 434–439.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110,* 88–125.

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.

Shillcock, R. C. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge, MA: MIT Press.

Tanenhaus, M. K. & Brown-Schmidt, S. (2008). Language processing in the natural world. In B. C. M. Moore, L. K. Tyler, & W. D. Marslen-Wilson (Eds.), *The perception of speech: from sound to meaning. Philosophical Transactions of the Royal Society B: Biological Sciences, 363,* 1105–1122.

Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 710–720.

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 274–296.

Appendix A

Stimuli Used in Experiment 1

Embedding position	Target picture	Superset distractor	Frequency	Subset distractor	Frequency	Subset and picture semantic category
Initial	Pig	Couch	9	Cow	23	Animal
	Girl	Boil	20	Boy	217	Gender
	Lock	Keen	26	Key	72	Household
	Hand	Toad	3	Toe	10	Human anatomy
	Foot	Thump	5	Thumb	23	Human anatomy
	Arm	Niece	5	Knee	29	Human anatomy
	Crayon	Pence	7	Pen	19	Office/craft utility
	Needle	Pinch	6	Pin	14	Office/craft utility
	Gong	Belt	21	Bell	39	Percussion instrument
	Pliers	Sword	13	Saw	388	Tool
	Train	Bust	6	Bus	65	Vehicle
	Car	Tramp	3	Tram	4	Vehicle
Final	Sheep	Scarf	8	Calf	10	Animal
	Fish	Kneel	3	Eel	4	Animal
	Mouse	Pram	5	Ram	9	Animal
	Shirt	Probe	6	Robe	9	Clothing
	Tomato	Scorn	7	Corn	25	Foodstuff
	Curtains	Shrug	4	Rug	12	Household
	Bowl	Span	7	Pan	29	Household
	Nose	Cheer	9	Ear	42	Human anatomy
	Leg	Charm	19	Arm	106	Human anatomy
	Brain	Slung	5	Lung	9	Human anatomy
	Bus	Scar	7	Car	276	Vehicle
	Sun	Drain	17	Rain	72	Weather

Appendix B

Stimuli Used in Experiment 2

Embedding position	Target picture	Superset distractor	Frequency	Subset distractor	Frequency	Subset and picture semantic category
Aligned	Shark	Selfish	12	Fish	175	Animal
	Owl	Acrobat	0	Bat	11	Animal
	Whale	Imbecile	1	Seal	13	Animal
	Onion	Unicorn	1	Corn	25	Foodstuff
	Bowl	Childish	13	Dish	22	Household
	Foot	Trainee	1	Knee	29	Human anatomy
	Hand	Plateau	6	Toe	10	Human anatomy
	Skull	Trombone	1	Bone	27	Human anatomy
	Leg	Nominee	1	Knee	29	Human anatomy
	Arm	Tomato	7	Toe	10	Human anatomy
	Brain	Motorist	3	Wrist	20	Human anatomy
	Gong	Decibel	1	Bell	39	Percussion instrument
Misaligned	Walrus	Squatter	2	Otter	9	Animal
	Dolphin	Genteel	3	Eel	4	Animal
	Horse	Diagram	8	Ram	9	Animal
	Eye	Adhere	1	Ear	42	Human anatomy
	Nose	Cashier	2	Ear	42	Human anatomy
	Finger	Bestow	1	Toe	10	Human anatomy
	Lips	Pioneer	6	Ear	42	Human anatomy
	Suitcase	Disperse	2	Purse	9	Personal storage
	Hammer	Thorax	1	Axe	6	Tool
	Bus	Restrain	5	Train	79	Vehicle
	Car	Constrain	1	Train	79	Vehicle
	Sun	Refrain	4	Rain	72	Weather
	Cloud	Membrane	2	Rain	72	Weather

Received January 22, 2008
Revision received December 4, 2008
Accepted March 12, 2009 ■