Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 57 (2007) 460–482
Journal of Memory and Language
www.elsevier.com/locate/jml

The tug of war between phonological, semantic and shape information in language-mediated visual search

Falk Huettig [a,b,*], James M. McQueen [b]

[a] Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium
[b] Max Planck Institute for Psycholinguistics, Post Box 310, 6500 AH Nijmegen, The Netherlands

Received 23 June 2006; revision received 6 February 2007
Available online 29 March 2007

Abstract

Experiments 1 and 2 examined the time-course of retrieval of phonological, visual-shape and semantic knowledge as Dutch participants listened to sentences and looked at displays of four pictures. Given a sentence with *beker*, 'beaker', for example, the display contained phonological (a beaver, *bever*), shape (a bobbin, *klos*), and semantic (a fork, *vork*) competitors. When the display appeared at sentence onset, fixations to phonological competitors preceded fixations to shape and semantic competitors. When display onset was 200 ms before (e.g.) *beker*, fixations were directed to shape and then semantic competitors, but not phonological competitors. In Experiments 3 and 4, displays contained the printed names of the previously-pictured entities; only phonological competitors were fixated preferentially. These findings suggest that retrieval of phonological, shape and semantic knowledge in the spoken-word and picture-recognition systems is cascaded, and that visual attention shifts are co-determined by the time-course of retrieval of all three knowledge types and by the nature of the information in the visual environment.

Keywords: Attention; Eye movements; Phonological representations; Semantic representations; Visual representations

Introduction

Phonological, semantic and visual-shape knowledge can be retrieved from long-term memory using either spoken or visual information. To understand how information from spoken language and vision interacts in determining behavior, however, it is necessary to establish how these different knowledge types are retrieved when someone is confronted simultaneously by speech and visual input. In particular, one has to specify the time-course of those retrieval operations. This was the goal of the present study.

Our focus was on the retrieval of phonological, semantic and visual-shape knowledge during spoken-word recognition, and the concurrent retrieval of these three types of knowledge from visual displays. Our investigations made use of what has become known as the visual-world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; see Henderson & Ferreira, 2004, for review). Research in this paradigm measures eye movements to visual stimuli in response to those stimuli and to concurrent spoken language. The paradigm provides closely time-locked and fine-grained measures of ongoing cognitive processing, in the form of fixations to different positions in the visual

* Corresponding author.
E-mail addresses: Falk.Huettig@mpi.nl (F. Huettig), James.McQueen@mpi.nl (J.M. McQueen).

0749-596X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2007.02.001

display over time. Three hypotheses have been proposed to explain this eye-movement behavior.

The phonological hypothesis

Pictures of objects with a name that begins in the same way as the word being heard accrue increased fixations. Allopenna, Magnuson, and Tanenhaus (1998) asked participants, for example, to "*Pick up the beaker*" in the context of a computer screen including a beaker, a beetle, a speaker, and a carriage. They found a phonological effect: eye movements to both the beaker and the beetle increased as the word *beaker* unfolded over time. Soon after the acoustic offset of *beaker*, however, looks to the beetle decreased as looks to the beaker continued to increase. As predicted by many theories of spoken-word recognition (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997), cohort competitors (i.e., words beginning in the same way as the word actually presented) appear to be considered temporarily as lexical hypotheses during word recognition. These findings provide evidence that fixations in the visual-world paradigm are driven by matches at the phonological level. Allopenna et al. indeed interpret these results in terms of the activation of phonological representations in the word-recognition system, and present simulations with the TRACE model (McClelland & Elman, 1986) in support of this interpretation. This is what we term the *phonological* hypothesis: phonological representations are retrieved on the basis of acoustic information and visual information (the names of the pictures in the display), and attentional shifts to pictures are made when there is a match between the representations retrieved from the two modalities.

Other phonological effects in visual-world studies include rhyme competitor effects (e.g., looks to a speaker as *beaker* is heard; Allopenna et al., 1998); modulation of cohort competitor effects by changes in fine-grained phonetic detail (Dahan, Magnuson, Tanenhaus, & Hogan, 2001b; McMurray, Tanenhaus, & Aslin, 2002; Shatzman & McQueen, 2006a), or by lexical frequency (Dahan, Magnuson, & Tanenhaus, 2001a); and evidence for the activation of words embedded in other words (Salverda, Dahan, & McQueen, 2003). These findings also support the phonological hypothesis.

The visual hypothesis

Cooper (1974) mentions that participants tended to fixate a picture of a snake when hearing the word *wormed* (in the context "*just as I had wormed my way on my stomach*"). This finding suggests that fixations in visual-world studies may be based on visual similarity. It could not be due to phonological similarity, but it might have been due to semantic similarity (or because participants mistook the snake for a worm). More recent studies, however, have examined this issue more systematically. Looks appear to be directed to visually related entities (i.e., objects related in shape) which are phonologically and semantically unrelated (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2004, in press). For instance, Huettig and Altmann (in press) found that participants shifted overt attention to a picture of a cable during the acoustic unfolding of the word *snake* (cable and snake have a similar global shape but are unrelated in phonological form and in meaning).

These visual form effects, however, could be contingent either on perceptual information (i.e., the perceived shape of the object) or on stored conceptual knowledge (i.e., knowledge about the typical shape of the object). Huettig and Altmann (2004) explored this issue by testing whether the mapping between language and visual input can be mediated by color relations. Color is an object property that allows the investigation of this question since conceptual attributes (the stored color knowledge about an object) and perceptual attributes (the perceived but non-diagnostic color of an object) can be dissociated. Participants saw four objects. The spoken sentence mentioned one of them (e.g., *frog*) or did not. When the object referred to in the sentence was not present, the display contained the picture of an object typically associated with the same color as the referent (e.g., *lettuce*, as both frogs and lettuces are associated with the color green). In one experiment, the picture of the lettuce was colored green; in another, no color was present. Participants' attention was drawn toward the competitor more than toward a distractor only when the competitor's color was visually present. In addition, it was found that participants, on hearing target words that are associated with a prototypical color, such as *pea*, looked towards a picture displayed in that color even though the referent of the picture was not itself associated with that color (e.g., a green blouse). In other words, on accessing the prototypical color information of the target referent, participants shifted overt attention to something with the same surface color. The perceived surface color rather than the stored color knowledge about a visual object thus seems to mediate overt attention. Knowledge about the visual features (e.g., color) of the object mentioned in the speech, however, must of course also be retrieved.

It therefore appears that match in terms of visual form can determine behavior in the visual-world paradigm. This, then, is the *visual* hypothesis. The probability of fixating a particular visual object reflects a match between stored knowledge of visual features (accessed by the spoken word) and visual features extracted from the display, namely, "a coarse structural representation associated with each of the object's locations" (Dahan & Tanenhaus, 2005, p. 457).

Demonstrations supporting the visual hypothesis (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2004, in press) cannot be explained in terms of phonological matches since phonological overlap was controlled. But, as Dahan and Tanenhaus (2005) argue, the reverse is not the case: apparent phonological effects could be explained by the visual hypothesis. If information processing in the speech recognition system is cascaded (and there is strong evidence that this is the case, see McQueen, Dahan, & Cutler, 2003), then the visual (and semantic) features of a beetle, for example, could be retrieved on the basis of the initial sounds of *beaker* alone (just like the visual and semantic features of the beaker itself). Participants may therefore look at phonological competitors not because of the phonological match between them and the current spoken word but because of a visual match. The process of retrieval of visual features would of course have to be mediated by phonological representations (e.g., the representation of the phonological form of *beetle* is a necessary link between the acoustic information and stored visual [and conceptual] knowledge about that word). According to the visual hypothesis, however, these phonological representations do not play a direct role in determining eye-movement behavior. One goal of the present research was thus to test if phonological effects can indeed be explained by the visual hypothesis, or, alternatively, if eye movements in the visual-world paradigm are determined by both phonological and visual matches.

The semantic hypothesis

There is also clear evidence, however, that language-mediated eye movements are sensitive to semantic relations (in the absence of any visual or phonological similarity) between visual objects and concurrent spoken words. Cooper (1974) found that participants were more likely to fixate pictures showing a snake, a zebra, or a lion when hearing a semantically-related word (*Africa*) than they were to fixate referents of semantically-unrelated control words. But Cooper failed to control the nature of the semantic similarity between the spoken words and the pictures. In addition, some of the items he used shared associative relationships. It is therefore unclear whether these effects were driven by conceptual similarity or by mere association.

Huettig and Altmann (2005) investigated whether semantic properties of individual lexical items can direct eye movements towards objects in the visual field in the absence of any associative relationships between the words heard and the concurrent visual objects (and indeed in the absence of any visual or phonological relationships). Visual displays containing four pictures of common objects were presented together with spoken sentences. Participants directed their overt attention towards a picture of an object such as a trumpet when a semantically related but non-associated word (e.g., *piano*) was heard. Three different measures of semantic relatedness (McRae feature norms, Cree & McRae, 2003; LSA, Landauer & Dumais, 1997; Contextual Similarity, McDonald, 2000) each separately correlated well with fixation behavior (Huettig & Altmann, 2005; Huettig, Quinlan, McDonald, & Altmann, 2006). Yee and Sedivy (2006) present similar evidence that eye movements can be determined by retrieval of semantic knowledge (even for a subset of items with no associative relationships).

It thus appears that eye movements in the visual-world paradigm can be driven by semantic matches. In other words, there is support for a third hypothesis about how behavior in this paradigm is determined. According to the *semantic* hypothesis (Huettig & Altmann, 2005; Huettig et al., 2006), the probability of fixating a particular visual object reflects the semantic similarity between conceptual knowledge accessed by the spoken word and the conceptual knowledge accessed from the visual objects. Furthermore, it is possible that phonological effects could be semantic effects in disguise. Participants could look at the beetle, for example, because its semantic features could be retrieved as the participant is hearing the beginning of *beaker*.

What kind of tug of war?

The evidence is already strong that there are independent shape and semantic effects in the visual-world paradigm (in the studies examining visual effects, semantic relationships were avoided, and vice versa, and in both cases phonological relationships were avoided). This evidence thus suggests that neither the simple visual or semantic hypotheses are correct, and that instead there is minimally what can be characterized as a two-way tug of war between fixations determined by visual-feature matches and those determined by semantic matches. As we have discussed, it is possible that phonological effects in the paradigm could be explained in terms of matches at either the visual-feature or semantic levels. Alternatively, however, the phonological hypothesis might also be true. If this were the case, then attentional shifts in the visual-world paradigm would be determined by a three-way tug of war among matches at phonological, visual-feature and semantic levels of representation.

We tested these alternative accounts by examining the time-course of retrieval of phonological, visual-shape and semantic knowledge as Dutch participants listened to sentences and looked at displays of four pictures on a computer screen. The display consisted of a phonological cohort competitor, a shape competitor, a semantic competitor, and an object unrelated on all three dimensions. If phonological effects in the visual-world paradigm are due to the matching of visual features, then there should be no difference between the time-course of fixations to the phonological competitors and that to the visual

competitors. Similarly, if phonological effects are due to the matching of semantic features, then the time-courses of looks to the phonological and semantic competitors should be the same. If, however, phonological effects reflect phonological matches, as the phonological hypothesis predicts, then looks to the phonological competitors should precede looks to the other competitors. This is because, as assumed in cascaded models of spoken-word recognition, retrieval of phonological representations should tend to precede retrieval of visual-feature and semantic representations.

The key prediction therefore concerned whether fixations to the phonological competitors would tend to be earlier than fixations to the other competitors. We could of course also examine the relative timing of fixations to visual and semantic competitors. Does retrieval of knowledge of the prototypical visual features of an entity associated with a spoken word precede, run concurrently with, or follow the retrieval of semantic knowledge such as category membership?

This question, and that about the relative timing of looks to phonological competitors, addresses not only whether information processing in the word-recognition system is cascaded, but also, more fundamentally, the extent to which representations of different types of knowledge (phonological, conceptual, visual) are independent. For example, if the phonological and semantic knowledge associated with a given word were stored in the same entry in the mental lexicon, such that retrieval of one type of knowledge necessarily entailed retrieval of all components of that entry, then there could be no difference in the time-course of phonological and semantic effects.

Exactly the same questions about flow of information and independence of representations that can be asked about word recognition can also be asked about picture recognition. This involves both visual and linguistic processing, since retrieval of phonology associated with pictures (i.e., the pictures' names) engages language production processes. Picture naming and picture–word interference studies have indeed been at the heart of chronometric analyses of speech production. In the Levelt, Roelofs, and Meyer (1999) theory of lexical access in production, picture names are only retrieved for selected words (''only selected lemmas will become phonologically activated'', p. 15). Thus, in contrast to the widely-held assumption that spoken-word recognition is cascaded, the assumption in this model is that spoken-word production is a serial process. There is considerable evidence (e.g., Levelt et al., 1991; see Levelt et al., 1999, for review) in favor of this view, though a number of more recent findings (e.g., Griffin & Bock, 1998; Morsella & Miozzo, 2002; Peterson & Savoy, 1998) suggest that there may be at least limited cascade in production.

We addressed further whether there is cascaded processing in picture recognition. If eye-movements are determined by matches at the level of visual features, then the components of this match based on input from vision are features that can be retrieved very readily from long-term memory, if not extracted directly from the information in the display (Huettig & Altmann, 2004). But if matches arise at a semantic level, then this depends on flow of information from stages of visual feature analysis to semantic levels. Similarly, phonological matches would depend on the further cascade of information to phonological levels. Thus, if Experiment 1 were to show that fixations to phonological competitors have a different time-course from fixations to the other types of competitors, then this would suggest cascade of information through the picture-recognition system as far as the retrieval of picture names.

The displays on experimental trials did not contain pictures corresponding to the spoken words. This was done in order to maximize the number of fixations to the competitors. On each of an equal number of filler trials, however, a picture of an object mentioned in the sentence was in the display. Hence a referent of a spoken word on a given trial was as likely to be present in the display as absent.

The participants' task was a passive one (to listen to the sentences and look at the display, but with no explicit instructions to search for particular targets). It was thus not necessary to include referents of words in all displays. Many previous visual-world studies have required participants to manipulate the objects that are mentioned (Tanenhaus et al., 1995; Allopenna et al., 1998). Other studies, however, have used listening-only tasks (Altmann & Kamide, 1999; Huettig & Altmann, 2005). An advantage of such tasks is that the paradigm can be extended beyond situations that require object manipulation. Effects obtained using listening-only tasks may suggest more general properties of the mapping between language and vision than effects that may be limited to goal-directed tasks. In any case, it appears that similar results can be obtained with active and passive versions of the paradigm. The data on semantic effects from active and passive tasks converge (compare Yee & Sedivy, 2006, with Huettig & Altmann, 2005), as do the data on visual effects (compare Dahan & Tanenhaus, 2005, with Huettig & Altmann, 2004, in press). Thus, even without explicit instruction, participants appear to search the visual display for matches to the information in the speech signal.

Experiment 1

Method

Participants

Thirty members of the subject panel of the MPI for Psycholinguistics, all native speakers of Dutch, were

paid for their participation. All participants had normal or corrected to normal vision.

Stimuli

Quintuples of words were selected for 40 experimental trials (see Appendix A). Each set consisted of a critical base word, three related words, and one unrelated word. Each critical word was placed in a neutral sentence (e.g., for *beker*, beaker, *Uiteindelijk keek ze naar de beker die voor haar stond*, 'Eventually she looked at the beaker that was in front of her'). The critical words were not predictable in these contexts (other examples of the preceding contexts, in translation, include 'He thought of a word that rhymed with …', 'He dreamt that night about a …', and 'She turned round and saw the …').

The four other words in each set were picturable, and their pictures were used in the visual displays. Each of these displays (see Fig. 1) contained a phonological cohort competitor that was unrelated in shape and semantics (e.g., a beaver, *bever*, was paired to the critical word *beker*), a shape competitor that was unrelated in phonology and semantics (e.g., a bobbin, *klos*), a semantic competitor that was unrelated in phonology and shape, and with more than a mere associative relationship to the critical word, such as membership of the same semantic category (e.g., a fork, *vork*), and an object unrelated on all three dimensions (e.g., an umbrella, *paraplu*). Norming studies (see below) confirmed the suitability of the materials on the shape and semantic dimensions. Phonological overlap was defined as follows. The first (or only) syllable of the critical word had the same consonantal onset and vowel nucleus as the first (or only) syllable of the name of the phonological competitor. Phonological overlap thus involved the first two phonemes in 37 pairs (e.g., the [be] of *beker* and *bever*), and the first three phonemes in the other three pairs. In temporal terms, average phonological overlap between critical words and their phonological competitors (i.e., the mean duration of the overlapping phonemes in the acoustic waveforms of the critical words, as measured with Praat [http://www.fon.hum.uva.nl/praat/]), was 192 ms. As shown in Table 1, the names of the four types of picture were matched for Celex word frequency ($F(3,156) = 1.11$, $p > .1$), and number of syllables ($F(3,156) = .26$, $p > .1$) and letters ($F(3,156) = 1.10$, $p > .1$).

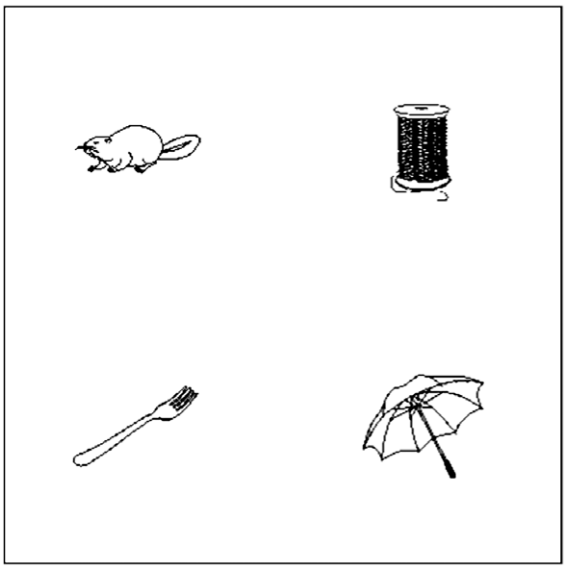

Fig. 1. Example of a visual stimulus used in Experiments 1 and 2. For the spoken sentence *Uiteindelijk keek ze naar de* ***beker*** *die voor haar stond*, 'Eventually she looked at the **beaker** that was in front of her', the display consisted of pictures of a beaver (the phonological competitor), a bobbin (the visual-shape competitor), a fork (the semantic competitor), and an umbrella (the unrelated distractor).

A further 40 quadruples of words were selected for filler trials. All four of the words in each set referred to picturable entities. These sets included one word that was placed in a neutral sentence context (like those used in experimental trials). The picture associated with that word, plus the three unrelated pictures associated with the other words from its set, were used in the visual displays.

Norming studies

Seven norming studies with Dutch native speakers were carried out using the Internet. None of the participants took part in any of the main experiments. In the first two studies participants rated either the shape or semantic similarity of the critical spoken words to the pictures in their matched displays. In the other five norming studies participants generated free associations (cf. Nelson, McEvoy, & Schreiber, 1998) for each of the targets and each of the display members in the experimental sets.

Similarity ratings. Thirteen participants provided shape-similarity ratings and 13 provided semantic-similarity ratings. In both studies participants were presented, over the Internet, with all 40 critical words, in printed form, each paired with their four pictures (displayed as in Experiment 1). Participants in Study 1 were asked to judge how similar the typical physical shape of the concept of the critical word was to the physical shape of the referents of the depicted objects while ignoring any similarity in meaning, using an 11-point scale (0 representing 'absolutely no similarity in physical shape' and 10 representing 'identical physical shape'). Participants in Study 2 used an 11-point scale to judge meaning similarity while ignoring shape similarity (0 representing 'absolutely no similarity in meaning', 10 representing 'identical meaning'). Results

Table 1
Properties of the materials

	Shape competitors	Semantic competitors	Phonological competitors	Unrelated distractors	Mean
Mean frequency (per million words)	18	15	33	33	25
Mean length in syllables	1.53	1.63	1.65	1.58	1.60
Mean length in letters	5.23	5.38	5.15	4.75	5.13

Table 2
Results of the similarity-rating norming studies: Means (and standard deviations in brackets) for the rated shape and meaning similarity between the critical words and each type of visual stimulus, and pairwise *t*-tests comparing stimulus types

		Shape competitors	Semantic competitors	Phonological competitors	Unrelated distractors
Shape similarity (Study 1)					
Mean rating (SD)		5.93 (0.68)	2.66 (0.62)	2.05 (0.55)	2.07 (0.52)
Comparisons with shape competitors	$t_1(1,12)$	—	17.9, $p < .001$	19.7, $p < .001$	19.3, $p < .001$
	$t_2(1,39)$	—	10.9, $p < .001$	14.8, $p < .001$	12.9, $p < .001$
Additional comparisons with unrelated distractors	$t_1(1,12)$	—	6.6, $p < .01$	< 1	—
	$t_2(1,39)$	—	2.6, $p < .05$	< 1	—
Meaning similarity (Study 2)					
Mean rating (SD)		2.78 (0.89)	5.93 (0.71)	2.06 (0.51)	1.80 (0.50)
Comparisons with semantic competitors	$t_1(1,12)$	8.7, $p < .001$	—	15.0, $p < .001$	16.0, $p < .001$
	$t_2(1,39)$	11.0, $p < .001$	—	16.2, $p < .001$	18.9, $p < .001$
Additional comparisons with unrelated distractors	$t_1(1,12)$	6.6, $p < .01$	—	4.1, $p < .05$	—
	$t_2(1,39)$	3.9, $p < .01$	—	1.9, $p > .05$	—

are shown in Table 2. Shape competitors were judged to be physically more similar to the critical words than any of the other visual stimuli, and semantic competitors were considered to be more similar in meaning to the critical words than any of the other stimuli.

Free word association. In Studies 3–7, participants saw a list of 40 words (again via the Internet) and were asked to type the first word that came to mind that was meaningfully related to each word. Separate association tests were carried out for the critical spoken words (Study 3), the shape (Study 4), semantic (Study 5), and phonological (Study 6) competitors, and the unrelated distractors (Study 7). All stimuli were presented as printed words (not pictures). Between 19 and 21 participants took part in each study. Responses matched another member of an experimental set only four times (i.e., 0.1% of trials). For example, the critical word *dokter*, doctor, was given in response to the semantic competitor *spuit*, syringe, by two participants. Results in the present experiments could therefore not be driven by simple associative relationships.

Procedure

The 40 experimental and 40 filler sentences were read aloud by a female native speaker of Dutch in a sound-damped booth. Digital recordings of these utterances, at a sample rate of 44.1 kHz with 16-bit resolution, were stored directly on computer. The sentences were read with a neutral intonation contour such that, in particular, the critical words were not highlighted. Pictures were black-on-white line drawings, selected from the MPI picture database.

Participants were seated at a comfortable distance from the computer screen. One centimeter on the visual display corresponded to approximately 1° of visual arc. The eye-tracking system was mounted and calibrated. Eye movements were monitored with an SMI Eyelink eye-tracking system, sampling at 250 Hz. Spoken sentences were presented to the participants through headphones. The parameters of each trial were as follows. First, a central fixation point appeared on the screen for 500 ms, followed by a blank screen for 600 ms. Then four pictures appeared on the screen as the auditory presentation of a sentence was initiated. The positions of the pictures were randomized across four fixed positions of a (virtual)

5 × 5 grid on every trial (grid positions 7, 9, 16 and 18 counting left to right and top to bottom; see Fig. 1).

Participants were not asked to perform any explicit task. They were told that they should listen to the sentences carefully, that they could look at whatever they wanted to, but that they should not take their eyes off the screen throughout the experiment (cf. Huettig & Altmann, 2005). Participants' fixations for the entire trial were thus completely unconstrained and participants were under no time pressure to perform any action.

Each participant was presented with all 80 trials. Experimental and filler trials were presented in random order. A central fixation point appeared on the screen after every five trials, allowing for some automatic drift correction in the calibration.

Data coding procedure

The data from each participant's right eye were analyzed and coded in terms of fixations, saccades, and blinks, using the algorithm provided in the Eyelink software. The timing of the fixations was established relative to the onset of the critical word in the spoken utterance. Graphical analysis software performed the mapping between the position of fixations and the pictures present on each trial, and displayed them simultaneously. Each fixation was represented by a dot associated with a number which denoted the order in which the fixation had occurred; the onset and duration of each fixation were available for each fixation dot. Fixations were coded as directed to the phonological competitor picture, the shape competitor picture, the semantic competitor picture or to the unrelated distractor picture, or to anywhere else on the screen. Fixations that fell within

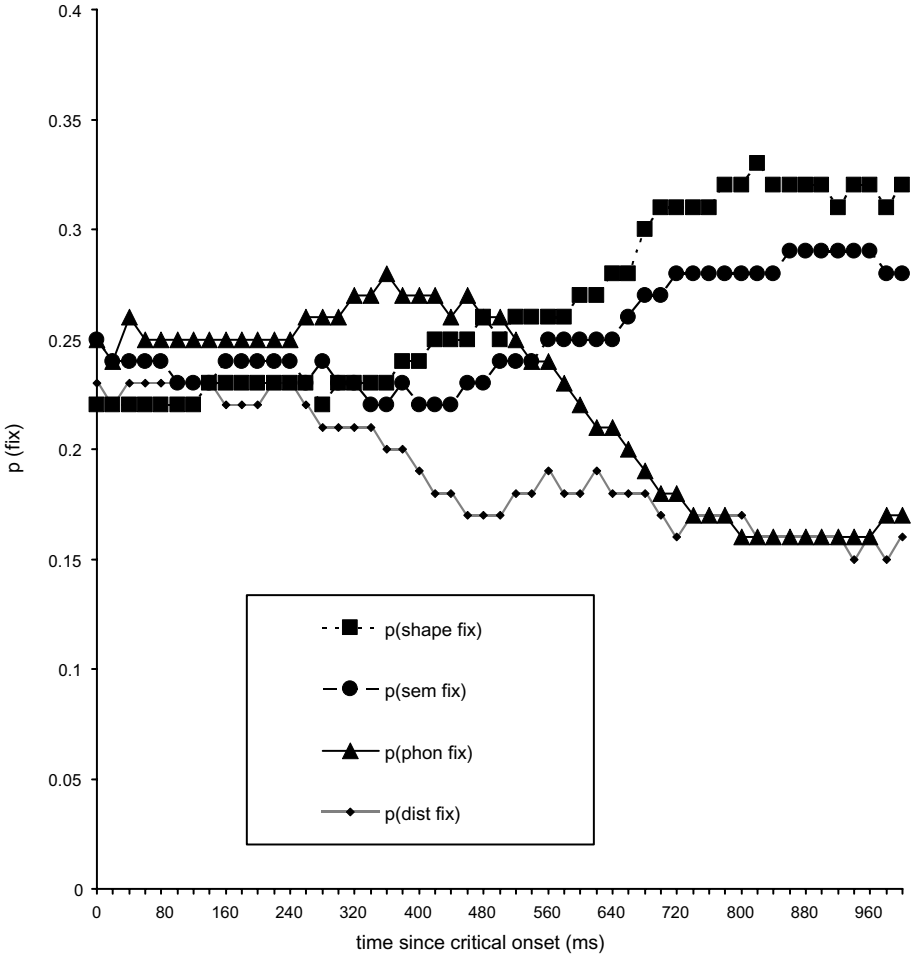

Fig. 2. Time-course graph showing fixation probabilities to phonological competitors, visual-shape competitors, semantic competitors, and unrelated distractors in Experiment 1 (picture displays, presented at sentence onset).

the cell of the grid in which a picture was presented were coded as fixations to that picture (cf. Salverda et al., 2003). The grid was only used during the coding procedure; it was not visible to participants during the experiment.

Results

Fig. 2 shows a time-course graph that illustrates the fixation proportions at 20 ms intervals to the various types of pictures over the course of the average experimental trial. *p*(phon fix) is the probability of a fixation to the phonological competitor, and was computed by counting, over all participants and items, the proportion of fixations to the phonological competitor in a given 20 ms time window relative to the total number of fixations in that time window. *p*(shape fix) and *p*(sem fix) are the corresponding fixation probabilities for the shape and semantic competitors respectively, and *p*(dist fix) is the corresponding fixation probability for the unrelated distractors. Zero represents the acoustic onset of the spoken critical word. The average fixation duration during the 1000 ms after critical word onset was 341 ms.

For the statistical analyses we computed mean fixation proportions for each type of picture over a time interval starting from 200 ms (as an estimate of the earliest point in time at which a fixation could reflect a response based on information in the critical word) and extending to 1000 ms. We calculated the ratio between the proportion of fixations to a particular competitor (phonological, shape, or semantic) and the sum of the particular competitor- and distractor-fixation proportions. We then compared the mean ratio (by participants and items) to .5 (cf. Dahan & Tanenhaus, 2005). A ratio greater than .5 shows that, of all the fixations directed toward a particular type of competitor and the unrelated distractors, the competitors attracted more than half of those fixations. Subsequent analyses examined in more detail the time-course of any effects that were significant in the overall analyses. We split the 200–1000 ms interval into eight 100 ms windows, and tested, for the data in each window, whether the competitor–distractor ratio, if it was significant overall, was significant in that window. These analyses provide estimates of when competitor and distractor fixation proportions diverge (and perhaps later converge) over the critical time interval.

As can be seen from Fig. 2, all types of pictures were fixated with an approximately equal probability at the acoustic onset of the critical word. However, as information from the critical word unfolded, *p*(phon fix), *p*(shape fix) and *p*(sem fix) diverged from *p*(dist fix). Importantly, *p*(phon fix) diverged from *p*(dist fix) about 240 ms after the onset of the critical word, whereas *p*(shape fix) and *p*(sem fix) diverged from *p*(dist fix) only about 100 ms later.

One-sample *t* tests showed that the phonological competitors [mean ratio of .55, $t_1(29) = 3.4$, $p = .002$, mean difference = .05, ±95% CI: .0277; $t_2(39) = 2.5$, $p = .017$], the shape competitors [mean ratio of .60, $t_1(29) = 8.7$, $p < .001$, mean difference = .10, ±95% CI: .0490; $t_2(39) = 4.2$, $p < .001$], and the semantic competitors [mean ratio of .59, $t_1(29) = 6.6$, $p < .001$, mean difference = .09, ±95% CI: .0532 $t_2(39) = 3.4$, $p = .002$] were fixated more than the unrelated distractors. There was no difference between fixations to the shape competitors and the semantic competitors [mean ratio of .52,

Table 3
Time-window analyses for Experiment 1

Time window from critical word onset (ms)	Phonological competitor ratio		Shape competitor ratio		Semantic competitor ratio	
	Mean	*t*-test	Mean	*t*-test	Mean	*t*-test
200–299	.54	$t_1 = 1.9$ ($p = .064$) $t_2 = 1.5$ ($p = .133$)	.51	$t_1 = .4$ ($p = .699$) $t_2 = .2$ ($p = .831$)	.52	$t_1 = 1.0$ ($p = .327$) $t_2 = .4$ ($p = .718$)
300–399	.57	$t_1 = 3.3$ ($p = .003$) $t_2 = 2.2$ ($p = .031$)	.53	$t_1 = 1.2$ ($p = .237$) $t_2 = 1.0$ ($p = .304$)	.53	$t_1 = 1.4$ ($p = .177$) $t_2 = .5$ ($p = .629$)
400–499	.59	$t_1 = 4.2$ ($p < .001$) $t_2 = 3.1$ ($p = .004$)	.58	$t_1 = 3.5$ ($p = .001$) $t_2 = 2.5$ ($p = .018$)	.56	$t_1 = 2.8$ ($p = .008$) $t_2 = 1.3$ ($p = .194$)
500–599	.57	$t_1 = 3.3$ ($p = .003$) $t_2 = 2.1$ ($p = .041$)	.59	$t_1 = 4.1$ ($p < .001$) $t_2 = 3.0$ ($p = .005$)	.58	$t_1 = 4.4$ ($p < .001$) $t_2 = 2.1$ ($p = .046$)
600–699	.53	$t_1 = 1.4$ ($p = .180$) $t_2 = 1.5$ ($p = .148$)	.61	$t_1 = 5.5$ ($p < .001$) $t_2 = 3.3$ ($p = .002$)	.59	$t_1 = 4.8$ ($p < .001$) $t_2 = 2.5$ ($p = .018$)
700–799	.51	$t_1 = .5$ ($p = .638$) $t_2 = 1.2$ ($p = .249$)	.65	$t_1 = 8.7$ ($p < .001$) $t_2 = 4.5$ ($p < .001$)	.63	$t_1 = 6.9$ ($p < .001$) $t_2 = 3.5$ ($p = .001$)
800–899	.50	$t_1 = -.2$ ($p = .914$) $t_2 = 1.2$ ($p = .244$)	.67	$t_1 = 10.5$ ($p < .001$) $t_2 = 5.1$ ($p < .001$)	.64	$t_1 = 6.5$ ($p < .001$) $t_2 = 4.0$ ($p < .001$)
900–999	.51	$t_1 = .5$ ($p = .559$) $t_2 = 1.5$ ($p = .144$)	.67	$t_1 = 10.1$ ($p < .001$) $t_2 = 5.3$ ($p < .001$)	.65	$t_1 = 6.8$ ($p < .001$) $t_2 = 4.4$ ($p < .001$)

$t_1(29) = 1.5$, $p > .1$, mean difference = .02, ±95% CI: .0365; $t_2(39) = .6$, $p > .1$].

Table 3 shows the time window analyses for the 100 ms time windows. During the 200–299 ms time window there were no significant differences. During the 300–399 ms time window, however, there were significantly more fixations directed to the phonological competitors than the unrelated distractors. No such differences were observed for the shape and semantic competitors. During the 400–499 ms and 500–599 ms time windows the phonological competitors, the shape competitors, and the semantic competitors were all fixated more than the unrelated distractors (though the effect was not significant by items for the semantic competitors in the 400–499 ms window). During the 600–699 ms time window there were no significant differences in looks to the phonological competitors and the unrelated distractors, but the shape and semantic competitors continued to accrue significantly more fixations than the unrelated distractors. This pattern of results remained the same for the remaining time windows.

Correlational analyses were carried out comparing the eye-movement data with the results of the similarity rating tasks. Specifically, the overall (200–1000 ms) differences in fixation probabilities between competitors and unrelated distractors for each item in each condition, as measures of the phonological, shape and semantic effects respectively, were compared to the similarity ratings for each item. There was a significant positive correlation between shape-similarity ratings and the shape effect ($r(40) = .573$, $p < .001$), indicating that the tendency to look at the shape competitor increased as the judged visual similarity of that competitor to the critical word increased. There was no such correlation of shape similarity with the phonological effect, and a negative correlation ($r(40) = -.358$, $p < .05$) with the semantic effect. This latter effect likely reflects the fact that there was also a strong negative correlation between the shape and semantic effects themselves ($r(40) = -.55$, $p < .001$). For an item set with a shape competitor that was highly physically similar to the critical word, it appears not only that this led to an increase in fixations to the shape competitor but also a corresponding decrease in fixations to the semantic competitor. Consistent with this account, none of the correlations of the semantic ratings with the eye-movement effects was significant.

Discussion

Experiment 1 revealed that attentional shifts to the phonological competitors preceded those to the visual-shape and semantic competitors. These results are consistent with the hypothesis that eye movements in the visual-world paradigm are the result of a three-way tug of war among fixations determined by matches at a phonological level of processing, those determined by matches of visual features, and those determined by semantic matches. If phonological effects were due to matches at the visual level (e.g., matching the visual features of a beaver, retrieved on the basis of partial phonological evidence of *bever*, given *beker* in the spoken sentence, to the features in the visual display), or, analogously, at the semantic level, then looks to the phonological competitors ought not to have had a time-course different from the time-courses of looks to the other types of competitor.

These results also suggest that information processing is cascaded. In the speech-recognition system, it appears that information flows continuously from the speech signal, via a phonological level of lexical representation, to levels of processing where knowledge about visual and semantic features can be retrieved and used. Fixations to phonological competitors, based on phonological matches, thus tended to precede fixations to the shape and semantic competitors. Note, however, that there was no reliable evidence of differential timing of retrieval of shape and semantic knowledge.

It appears that there is also cascade in the picture-recognition system, from the visual display to visual-feature representations, and from there to semantic representations and then to phonological representations. According to a strictly serial model of speech production, such as that of Levelt et al. (1999), the names of pictures should only be retrieved for selected lemmas (i.e., only those the participant in a picture-naming study intends to say). The present data challenge this view: Even though participants did not have to produce overtly any of the picture names, they still retrieved those names, as the fixations to the phonological competitors reveal.

There was ample time in Experiment 1 for picture names to be retrieved since the visual displays were present from the onset of the sentences and the critical words appeared, on average, after 6.85 other words. If the amount of time to process the visual display were reduced, then there might not be enough time for picture names to be retrieved, and attentional shifts towards the phonological competitors would be reduced or even eliminated. Experiment 2 explored this prediction.

Experiment 2

Experiment 2 was identical to Experiment 1 except that the visual display was presented 200 ms before the acoustic onset of the critical word rather than at sentence onset. We predicted that under these conditions overt attention would be driven primarily by matches of visual features and thus that fixations to the shape competitors would predominate. We further reasoned that fixations to semantic competitors would be delayed

relative to those to the shape competitors, since more time would tend to be required to retrieve semantic than visual-feature knowledge from the display. We also predicted that there would be little difference between the proportion of looks to the phonological competitors and those to the unrelated distractors. Experiment 1 has shown that the information in the speech signal tends to cause early fixations to phonological competitors—for the simple reason that, at later moments in time, the phonological competitors become inconsistent with the material in the speech signal. Our measurements have shown that the average amount of time before this point of inconsistency was about 190 ms. Levelt et al. (1991) estimate 385 ms for retrieval of the name of a single picture. So, given the 200 ms picture preview, name retrieval should tend to occur approximately at the point in time where the critical word diverges phonologically from the competitor name. We therefore assumed that, in Experiment 2, by the time picture names could be retrieved from the visual information, the phonological competitors would already be inconsistent with the speech material, and hence that there would be few, if any, looks to the phonological competitors.

Experiment 2 also addressed two alternative explanations for the results of Experiment 1. One could argue that the phonological effect in Experiment 1 was, after all, driven by visual feature matches. Consider the example again. When the participant has heard only the first two sounds of *beker*, *bever* and *beker* are more or less equally consistent with the speech signal. Retrieval of the visual features associated with *bever* and *beker* therefore ought to be more or less equally well advanced. In terms of the visual display, however, there is more evidence for a beaver than for a beaker. In particular, although a bobbin shares some visual features with a beaker, it is not a beaker. It is thus possible that fixations to the beaker precede those to the bobbin because, until the speech signal tells the participant that they are not listening to *bever*, the match between language and vision, in terms of visual features, is stronger for the beaver than for the bobbin. A parallel argument could be told with respect to matches at the semantic level.

If either of these accounts of Experiment 1 is correct, then looks to the phonological competitors should once again precede looks to the shape and semantic competitors in Experiment 2. That is, if this effect reflects differences in the strength of matches of visual (or semantic) features, then changing the timing of the presentation of the visual display should not alter the effect. The beaver, for example, should still be momentarily a better visual match to the first syllable of *beker* than the bobbin, because that syllable is also the first syllable of *bever*. For the same reason, the beaver should also remain a better semantic match to the first syllable of *beker* than the fork.

Method

Participants

Thirty members of the MPI for Psycholinguistics subject panel, all native speakers of Dutch, were paid for their participation. None had participated in Experiment 1. All participants had normal or corrected to normal vision.

Stimuli and procedure

The same stimuli as in Experiment 1 were used. The procedure was also identical to the earlier experiment except that the visual display appeared only 200 ms before the acoustic onset of the critical word.

Results

Fig. 3 shows the fixation proportions at 20 ms intervals to the various types of pictures over the course of the average trial, as in Fig. 2.

A one-sample t test revealed that the shape competitors were fixated more than the unrelated distractors [mean ratio of .59, $t_1(29) = 7.3$, $p < .001$, mean difference = .09, ±95% CI: .0249; $t_2(39) = 3.8$, $p < .001$]. The shape competitors were also reliably more fixated than the semantic competitors [mean ratio of .56, $t_1(29) = 7.2$, $p < .001$, mean difference = .06, ±95% CI: .0178; $t_2(39) = 3.2$, $p < .01$]. Differences in performance between the semantic competitors and the unrelated distractors were much less reliable [approaching significance by participants but not by items, mean ratio of .527, $t_1(29) = 2.0$, $p = .059$, mean difference = .027, ±95% CI: .0282; $t_2(39) = .9$, $p > .1$]. There were no differences in overt attention between the phonological competitors and the unrelated distractors [mean ratio of .506, $t_1(29) = .7$, $p > .1$, mean difference = .006, ±95% CI: .018; $t_2(39) = -.1$, $p > .1$].

Table 4 shows the time window analyses for the shape and semantic competitor conditions. During the first three time windows there were no significant differences. During the 500–599 ms time window, however, there were more fixations (significant by participants only) directed to the shape competitors than the unrelated distractors. No such differences were observed between the semantic competitors and the unrelated distractors. During the 600–699 ms time window the shape competitors (but not the semantic competitors) were fixated significantly more often than the unrelated distractors. This increased attention to the shape competitors remained the same for the subsequent time windows. During the 700–799 ms time we also observed a tendency to look more at the semantic competitors than the unrelated distractors. During the subsequent time windows the semantic competitors accrued significantly more fixations than the unrelated distractors.

Correlations of similarity ratings with eye-movement behavior, like those in Experiment 1, were also carried

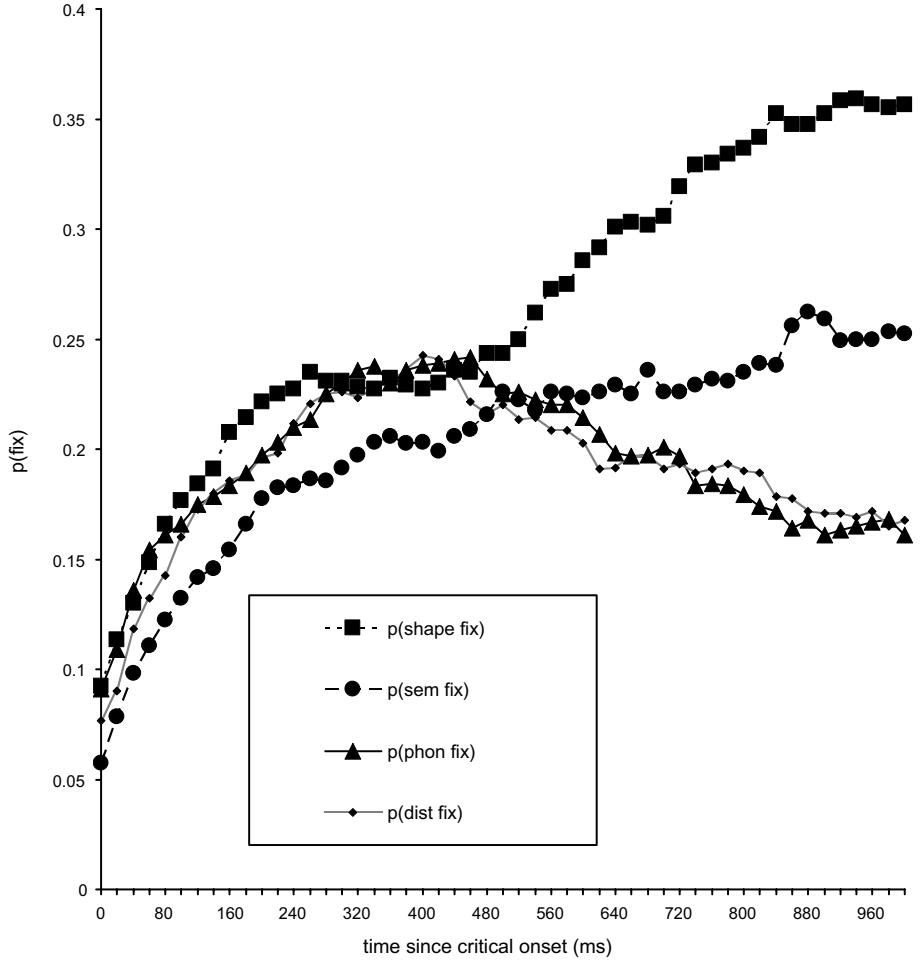

Fig. 3. Time-course graph showing fixation probabilities to phonological competitors, visual-shape competitors, semantic competitors, and unrelated distractors in Experiment 2 (picture displays, presented 200 ms before critical word onset).

out. The only effect involving the shape ratings was a correlation with the shape effect ($r(40) = .283$, $p = .077$). As in Experiment 1, the semantic ratings did not correlate with the eye-movement data.

Discussion

In Experiment 1, we observed early attentional shifts towards the phonological competitors. We argued that this finding depended in part on cascade of information through the picture-recognition system as far as the retrieval of picture names. The results of Experiment 2 support this view. Experiment 2 was identical to Experiment 1 except that only 200 ms preview of the visual display was given before the onset of the critical word. Under these conditions we observed no differences in overt attention between phonological competitors and unrelated distractors, presumably because there was not enough time for the retrieval of picture names before the evidence in the speech signal was able to rule out the phonological competitor as a viable lexical hypothesis in the word-recognition system.

A possible alternative explanation for the results of Experiment 1 was that the early looks to the phonological competitors reflected a better match, in terms of visual features rather than phonological form, between the knowledge retrieved on the basis of the speech signal and the visual representation of the phonological competitor than the match between the knowledge retrieved from speech and the shape competitor (e.g., the [be] of *beker* provides a better visual match to the beaver than to the bobbin because the beaver is indeed a beaver but the bobbin only shares some visual features with a beaker). This explanation predicts that the same early looks to phonological competitors should have been found in Experiment 2, since the opportunity for matches in terms of visual features was the same in the two experiments. Yet another explanation for the Exper-

Table 4
Time-window analyses for Experiment 2

Time window from critical word onset (ms)	Shape competitor ratio		Semantic competitor ratio	
	Mean	*t*-test	Mean	*t*-test
200–299	.52	$t_1 = 1.7$ ($p = .097$) $t_2 = .1$ ($p = .987$)	.46	$t_1 = -2.1$ ($p = .050$) $t_2 = -.3$ ($p = .794$)
300–399	.51	$t_1 = .3$ ($p = .787$) $t_2 = -.2$ ($p = .836$)	.47	$t_1 = -1.6$ ($p = .130$) $t_2 = -.6$ ($p = .557$)
400–499	.50	$t_1 = .3$ ($p = .789$) $t_2 = -.1$ ($p = .989$)	.47	$t_1 = -1.6$ ($p = .128$) $t_2 = -1.0$ ($p = .326$)
500–599	.55	$t_1 = 2.4$ ($p = .022$) $t_2 = 1.6$ ($p = .114$)	.51	$t_1 = .7$ ($p = .498$) $t_2 = .02$ ($p = .981$)
600–699	.60	$t_1 = 4.2$ ($p = .001$) $t_2 = 3.3$ ($p = .002$)	.54	$t_1 = 2.0$ ($p = .055$) $t_2 = 1.1$ ($p = .267$)
700–799	.63	$t_1 = 5.6$ ($p < .001$) $t_2 = 3.9$ ($p < .001$)	.55	$t_1 = 2.1$ ($p = .045$) $t_2 = 1.7$ ($p = .096$)
800–899	.66	$t_1 = 8.6$ ($p < .001$) $t_2 = 4.8$ ($p < .001$)	.57	$t_1 = 3.1$ ($p = .004$) $t_2 = 2.4$ ($p = .022$)
900–999	.70	$t_1 = 13.3$ ($p < .001$) $t_2 = 5.4$ ($p < .001$)	.62	$t_1 = 5.3$ ($p < .001$) $t_2 = 2.8$ ($p = .008$)

iment 1 results was a parallel one: that there was momentarily a better semantic match of the phonological competitor to the speech material than of the semantic competitor (e.g., a beaver is a beaver, but a fork only shares some semantic features with a beaker). The absence of a tendency to fixate the phonological competitor in Experiment 2 is inconsistent with both of these explanations.

One could argue, however, that because of the limited preview in Experiment 2, the influence of visual feature information from the visual display was delayed until after the limited time window in which there are phonological effects. On this account, the results of Experiments 1 and 2 would still be consistent with the visual hypothesis of Dahan & Tanenhaus (2005). We consider this account to be extremely unlikely, for three reasons. First, the previous literature suggests that, for the type of displays used here, information from those displays will start to be extracted at display onset and used immediately to modulate eye movements. A conservative estimate of the functional field of view for complex scenes is 4° (Henderson & Ferreira, 2004). For simple line drawings of four spatially distinct objects, such as these used here, functional field of view is likely to be much larger (Henderson & Ferreira, 2004; Parker, 1978). Pollatsek, Rayner, and Collins (1984), for instance, showed that line drawings of objects in an uncluttered field can be identified 10° from fixation. There was approximately 5.3° between the fixation cross and the centers of the pictures in our displays. The functional field of view at display onset therefore almost certainly included all four objects, and, if so, visual features could be extracted even before critical word onset.

The second reason we question this alternative account is that it appears inconsistent with other very similar visual-world data. The account requires that visual features, even if they start to be extracted from the display prior to critical word onset, cannot influence eye movements substantially until about 600 ms later. Experiment 1 shows that there is a delay of approximately 350 ms between the onset of acoustic information and when that information produces a statistically significant effect in the eye-movement record. Approximately 200 ms after that, therefore, the mismatching phonetic information (between the word actually spoken and the phonological competitor) should start to have an effect (the disappearance of a statistically significant difference), as indeed was found in Experiment 1 (the phonological effect was not significant in the 600–700 ms window). According to the visual hypothesis, this reflects the time-course of the availability of shape information corresponding to candidate spoken-word hypotheses. Since listeners in Experiment 2 were hearing exactly the same sentences as in Experiment 1, it is reasonable to suppose that the time-course of speech-based processing was also the same. Therefore, if use of visual features based on pictorial information was delayed in Experiment 2, then it would have to be delayed by about 600 ms to remove fully the effect on the phonological competitors. This is unlikely, given that Dahan and Tanenhaus (2005) show, with a 300-ms display preview, that fixations to shape competitors start to diverge from those to unrelated distractors 300 ms after critical word onset.

Third, this account is inconsistent with our own data. One might argue that the comparison with the Dahan and Tanenhaus (2005) study is inconclusive, given differences in materials and preview times (200 vs. 300 ms). We can compare the shape and phonological effects in Experiments 1 and 2 directly, however. It appears that there is a global delay in effects in Experiment 2, perhaps caused by the onset of the visual display at unpredictable

times in the middle of the spoken sentences (a similar delay was observed by Dahan and Tanenhaus, who compared 300 ms with 1000 ms preview). But the effect for shape competitors in Experiment 2 is nonetheless present in the 500–600 ms window and fully significant in the 600–700 ms window. If the visual features of the shape competitors can influence fixations around 600 ms after word onset, then certainly so should the visual features of the phonological competitors (recall that the visual feature overlap is stronger for the phonological competitors than for the shape competitors). This is about the earliest time at which the phonological mismatch could start to have a reliable effect. Thus, even if, in contradiction of the literature, information from the visual display only starts to influence fixations about 800 ms after display onset, one would have to predict at least some preferential fixations to the phonological competitor around this time too (or earlier, as in Experiment 1), if those fixations are driven by visual matches.

These three arguments suggest that the phonological effect in Experiment 1 is not due to matching of visual features.[1] The idea behind the manipulation of display timing in Experiment 2 was that briefer preview would delay retrieval of the picture names relative to the temporal unfolding of the speech signal, and thus remove the phonological effect. The manipulation was successful, we argue, because of the additional processing time required for picture name retrieval. We predict that a larger manipulation (e.g., onset of the display delayed until after acoustic onset of the critical word) would be required to remove effects based on visual match.

Experiment 2 thus confirms the hypothesis of a three-way tug of war among phonological, visual-feature and semantic matches in language-mediated visual search. According to this view, the tug of war changes over time, with different types of matches tending to dominate behavior as processing of both spoken and visual material unfolds. Experiment 2 provided additional support for the hypothesis that information processing is cascaded in the picture-recognition system. Unlike in Experiment 1, shifts to the shape competitors preceded shifts to the semantic competitors. If picture recognition involves continuous flow of information from visual processing levels to semantic levels, visual features should be available earlier for word–picture matching than semantic features. With only limited preview of the display in Experiment 2 this timing effect could be observed; in Experiment 1, however, there was enough time for picture recognition to advance even beyond the semantic level (i.e., to the picture name) before the critical spoken word was heard.

Experiment 3

Experiments 1 and 2 suggest that language-mediated attentional shifts in the visual-world paradigm are co-determined by the type of information in the visual display, by the temporal unfolding of information in the speech signal and by the timing of processing within both the picture- and word-recognition systems. In particular, these results suggest that behavior in the visual-world paradigm cannot be explained in terms of word–picture matching at the level of visual features alone, nor at the level of semantic features alone, but that it involves matching in terms of visual, semantic and phonological features. If eye movements to pictures can be determined, in part, by phonological matches, then phonological effects should be stronger if the display contains printed words. We examined this possibility in Experiment 3.

The pictures used in the earlier experiments were replaced with printed words. Retrieval of phonological knowledge given a picture depends on both visual-feature processing and semantic processing. Reading a word provides much more direct access to phonological knowledge. Although models of reading differ substan-

[1] Mike Tanenhaus has pointed out that we cannot rule out the possibility that signal-driven fixations in Experiment 2 did not begin until after the phonetic information was inconsistent with the phonological competitor. If signal-driven fixations did not begin until after this information could determine eye-movement behavior, then this could explain the absence of a phonological effect in Experiment 2, and the account could thus be maintained that the phonological effect in Experiment 1 was due to matching of visual features. Looks to the shape competitors started to increase around 500 ms after critical word onset, and those to the unrelated distractors started to decrease about 60 ms earlier. There was thus no evidence of signal-driven fixations until about 250 ms after the average point where the phonological competitors became inconsistent with the input (190 ms after word onset). The issue, however, is not when the inconsistent phonetic information arrives, but when that information determines behavior. The results of Experiment 1 suggest that it took at least 600 ms from word onset for the proportion of fixations to the phonological competitors to drop to the level of the unrelated distractors. The longest estimate of the time window in which phonological effects can be observed is thus that it extends for 400 ms after the point of phonetic inconsistency, which would then partially overlap with when signal-driven fixations were observed in Experiment 2. It is possible, however, that the actual window is somewhat shorter than this estimate. Although we consider it unlikely that the lack of a phonological effect in Experiment 2 reflects a timing mismatch (i.e., no signal-driven fixations before the phonological competitor ceased to be able to attract fixations) we cannot yet exclude this possibility. It is clear that interpretation of the results of Experiments 1 and 2 depends on the precise timing of information processing in both the visual and spoken modalities. Further research will be required to specify in greater detail the time-course of the uptake and use of information from visual displays and from the speech signal.

tially in their assumptions about the mechanisms and representations involved, they agree that word phonology can be retrieved either via semantic representations or more directly from the orthographic input (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996). It has even been proposed that reading print for meaning requires phonological mediation (e.g., Van Orden, Johnston, & Hale, 1988). We therefore predicted that the use of printed words would cause an increase in the proportion of looks to phonological competitors, and would allow those fixations to emerge earlier in time. In particular, we predicted that if we used the same timing parameters as in Experiment 2, looks to the names of phonological competitors would be found. With only 200 ms of preview of pictures, we have suggested that there is not enough time for retrieval of picture names. With the same limited preview time, but with an array of printed words, we predicted that phonological knowledge could be retrieved, and hence that looks to phonological competitors would be observed. Experiment 3 was thus a repeat of Experiment 2, with the names of the same competitors appearing in the visual display 200 ms before the onset of the critical spoken words.

One can also ask what the fate of the shape and semantic competitors might be with a display of printed words. Following the previous logic, one might expect that eye movements to these types of competitor ought to be delayed relative to eye movements towards phonological competitors: Retrieval of phonological knowledge should be the fastest of the three types, in both spoken and visual word recognition. If anything, one would also expect looks to semantic competitors to precede looks to shape competitors. Experiments 1 and 2 show, however, that there is enough time for the retrieval of shape and semantic knowledge from the information in the speech signal to influence eye movements soon after the onset of the critical word. Delay in preferential fixations to semantic and/or shape competitors in Experiment 3 would therefore have to reflect insufficient time for the retrieval of these types of knowledge from the visual input. But semantic and especially pictorial visual features are arguably not relevant when the visual search is over an array of printed words. The participant might thus decide that matching features at these levels of representation is inappropriate with a printed-word display. An alternative outcome of Experiment 3, therefore, might be that participants only look preferentially at the phonological competitors.

Method

Participants

Twenty eight members of the MPI for Psycholinguistics subject panel, all native speakers of Dutch, were paid to take part. All participants had normal or corrected to normal vision. None had participated in the earlier experiments.

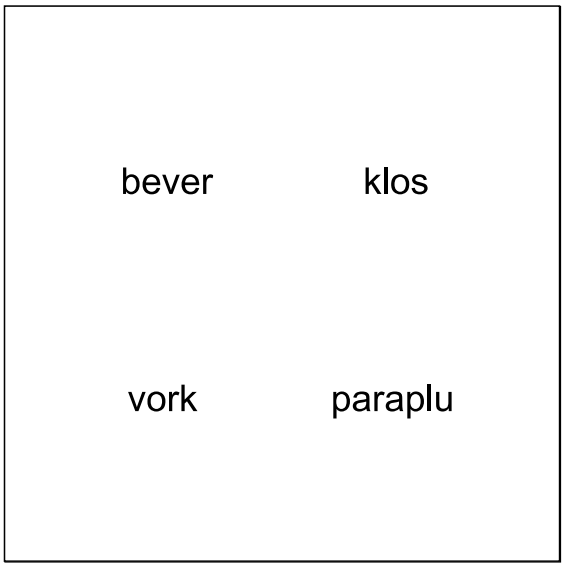

Fig. 4. Example of a visual stimulus used in Experiments 3 and 4. For the spoken sentence *Uiteindelijk keek ze naar de **beker** die voor haar stond*, 'Eventually she looked at the **beaker** that was in front of her', the display consisted of the printed words *bever* ('beaver', the phonological competitor), *klos* ('bobbin', the visual-shape competitor), *vork* ('fork', the semantic competitor), and *paraplu* ('umbrella', the unrelated distractor).

Stimuli

The same stimuli as in Experiments 1 and 2 were used but, instead of the visual objects, their printed Dutch names were presented (see Fig. 4). The words were presented in Arial font, centered in the same positions of the same virtual 5×5 grid that was used in Experiments 1 and 2.

Procedure

The procedure was the same as in Experiment 2. That is, the display of words appeared only 200 ms before the critical spoken word.

Results

Fig. 5 plots the data in the same way as in the earlier graphs. This figure shows that, at the acoustic onset of the critical word, all types of pictures were fixated with an approximately equal probability. However, as information from the critical word unfolds, only p(phon fix) diverges reliably from p(dist fix).

A one-sample t test revealed that the phonological competitors were fixated more than the unrelated distractors [mean ratio of .58, $t_1(27) = 6.2$, $p < .001$, mean

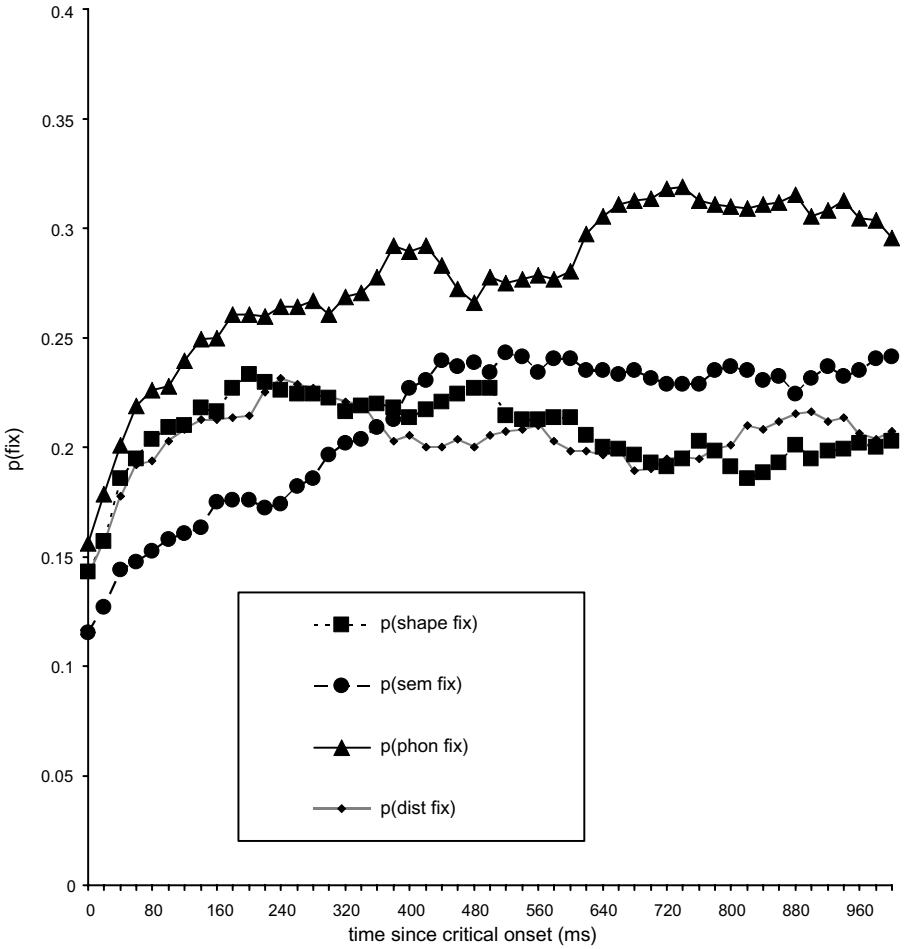

Fig. 5. Time-course graph showing fixation probabilities to phonological competitors, visual-shape competitors, semantic competitors, and unrelated distractors in Experiment 3 (printed word displays, presented 200 ms before critical word onset).

difference = .08, ±95% CI: .0273; $t_2(39) = 3.9$, $p < .001$]. The shape competitors [mean ratio of .50, $t_1(27) = -.1$, $p > .1$, mean difference = .0, ±95% CI: .0161; $t_2(39) = -.2$, $p > .1$] were not fixated more than the unrelated distractors. Performance between the semantic competitors and the unrelated distractors differed slightly [significant by participants but not by items, mean ratio of .52, $t_1(27) = 2.6$, $p = .013$, mean difference = .02, ±95% CI: .0155; $t_2(39) = .9$, $p > .1$].

We analyzed the time-course for the reliable phonological effect only. The fixations to the phonological competitor words first diverged significantly from those to the unrelated words in the 600–699 ms time window [mean ratio of .60, $t_1(27) = 5.0$, $p < .001$; $t_2(39) = 3.8$, $p < .01$].

Discussion

Experiment 3 revealed a large phonological effect but no attentional shifts to semantic or visual-shape competitors. With exactly the same duration of preview, and exactly the same speech input, there was either no fixation preference for phonological competitors when they were pictures (Experiment 2), or only a fixation preference for these competitors when they were printed words (Experiment 3). Hence phonological effects are indeed stronger with a printed-word display than with a display of pictures. This is consistent with the hypothesis that retrieval of phonological knowledge takes less processing time during visual-word recognition than during picture recognition.

We suggested earlier that the absence of preferential looks to shape and semantic competitors could reflect the irrelevance of this type of knowledge when the display consists of printed words. Semantic knowledge, and especially knowledge of pictorial features, may not be considered relevant by the participant when the visual search is over an array of ortho-

graphic word forms. The failure to observe fixations based on shape or semantic matches, however, could simply reflect delays in information processing. Just as the participants in Experiment 2 may have failed to look at the phonological competitors because picture names were not available early enough, it is possible that the participants in Experiment 3 did not look preferentially at the shape and semantic competitors because there was not enough time to retrieve visual and conceptual knowledge before the information in the speech signal made clear what the critical word was.

This timing account is somewhat implausible, since looks to visual and semantic competitors in Experiments 1 and 2 continued even after the acoustic offset of the critical spoken word. Fixations to these competitors could therefore have been delayed in Experiment 3, but are unlikely to have been ruled out entirely on the basis of limitations in processing time. Nevertheless, it was straightforward to examine this timing account further. In Experiment 4 we used the same printed-word materials as in Experiment 3, but used the timing parameters from Experiment 1. If participants are able to see the visual display from the onset of the spoken sentence, then there is surely enough time for the retrieval of both shape and semantic features from the words in the display before the onset of the critical spoken word. If the results of Experiment 4 replicate Experiment 3, then this would suggest that the preference for phonological competitors alone is not due to restrictions in processing time, but rather to the relevance of phonological knowledge, as opposed to the irrelevance of shape and semantic knowledge, when the visual environment consists of an array of printed words.

Experiment 4

Method

Participants

Thirty members of the MPI for Psycholinguistics subject panel, all native speakers of Dutch, were paid for their participation. All participants had normal or corrected to normal vision. None had taken part in the previous experiments.

Stimuli and procedure

The same stimuli as in Experiment 3 were used. The procedure was identical to Experiment 1. Critically, displays were presented at the onset of the spoken sentences.

Results

The data were again plotted in the same way. Fig. 6 shows that overall performance was similar to Experiment 3. As information from the critical word unfolds, p(phon fix) diverges from p(dist fix) and the phonological competitors then receive most fixations (though the function rises faster than in the previous experiment, presumably because of the increased preview).

A one-sample t test revealed that the phonological competitors were fixated more than the unrelated distractors [mean ratio of .59, $t_1(29) = 5.3$, $p < .001$, mean difference = .09, ±95% CI: .0341; $t_2(39) = 5.5$, $p < .001$]. The semantic competitors [mean ratio of .50, $t_1(29) = -.2$, $p > .1$, mean difference = .0, ±95% CI: .0265; $t_2(39) = -.1$, $p > .1$] were not fixated more than the unrelated distractors. The shape competitors accrued slightly less fixations (significant by participants but not by items) than the unrelated distractors [mean ratio of .47, $t_1(29) = -2.5$, $p = .018$, mean difference = .03, ±95% CI: .0280; $t_2(39) = -1.8$, $p > .05$].

We analyzed the time-course for the reliable phonological effect only. The fixations to the phonological competitor words first diverged significantly from those to the unrelated words in the 300–399 ms time window [mean ratio of .57, $t_1(29) = 3.0$, $p < .01$; $t_2(39) = 3.8$, $p < .01$].

Discussion

Experiment 4 was identical to Experiment 3 except that more preview of the visual display was given before the onset of the spoken critical word. The overall pattern of results replicates Experiment 3. There was a large shift in attention towards the phonological competitors (though this shift was faster than in Experiment 3 because of the greatly increased preview). There were no reliable shifts in attention towards the visual-shape and semantic competitors. Since there was ample time for the retrieval of shape and semantic knowledge based on the information in the visual display (and certainly for retrieval of these types of knowledge from the speech material, given the results of Experiments 1 and 2), we conclude that participants chose not to use these types of knowledge when they were confronted with printed-word displays. Therefore, unlike with displays of pictures, where there is a tug of war among fixations determined by speech–vision matches at all three levels of representation, matches between spoken and printed words appear to be predominately phonological.

Note that this account may also explain why participants continued to look at the phonological competitor in Experiments 3 and 4, even after the speech signal had indicated that this competitor was not the word the speaker intended. If phonological matches were all that was relevant, then even though the speech signal clearly indicates that the phonological competitor is not in the spoken sentence, there would be no other choices partic-

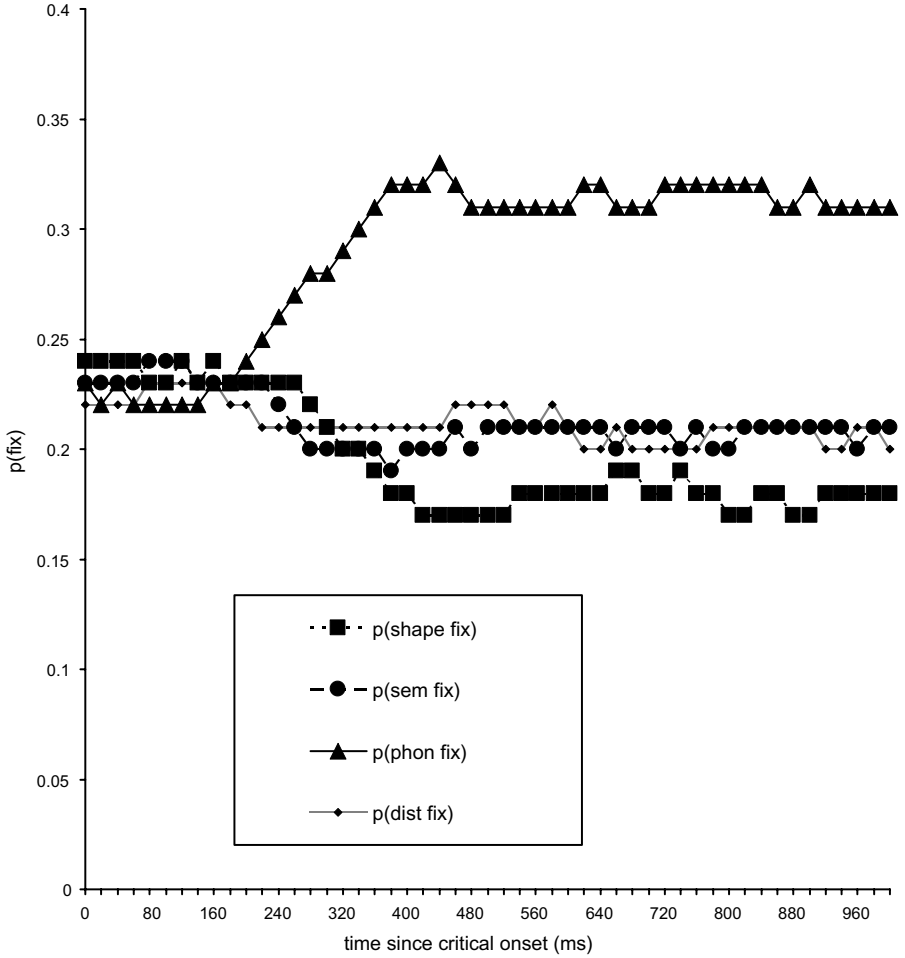

Fig. 6. Time-course graph showing fixation probabilities to phonological competitors, visual-shape competitors, semantic competitors, and unrelated distractors in Experiment 4 (printed word displays, presented at sentence onset).

ipants could turn their attention towards. Presumably they had read all four words in Experiment 4 before the onset of the critical spoken word. They were thus in little doubt that, after looking at the phonological competitor, there were no other (even partial) phonological matches in the display. It is reasonable to assume that participants were less certain of the content of the briefly previewed display in Experiment 3. This may explain why the bias towards the phonological competitor was somewhat weaker in Experiment 3 than in Experiment 4.

General discussion

In four experiments Dutch participants listened to spoken Dutch sentences while looking at visual displays. Even though the spoken sentences were identical across all experiments and even though the visual displays had exactly the same abstract content in each experiment, eye movement behavior, both in terms of where participants looked and when they looked, was radically different across the four experiments. All that changed was the relative timing of presentation of the linguistic and visual information, and whether the displays consisted of pictures or printed words. When participants had time to look at a display of four pictures from the onset of the sentences (Experiment 1), attentional shifts to phonological competitors of the critical spoken words preceded attentional shifts to shape competitors and to semantic competitors (that had more in common with the critical words than mere associative relationships). With only 200 ms of preview of the same picture displays prior to onset of the critical word (Experiment 2), participants did not look preferentially at the phonological competitors, and instead made more fixations to the shape competitors and then the semantic competitors. A third pattern was found when the pictures were

replaced with printed words (the names of the same entities as before). Now attentional shifts were made only to the phonological competitors, both when there was only 200 ms of preview (Experiment 3) and when the displays appeared at sentence onset (Experiment 4).

The visual-world paradigm

This study shows that it is possible to use printed words instead of pictures in the visual-world paradigm. This could be a real boon, since it frees the experimenter from the quite severe restriction that all critical materials have to be picturable. But our results suggest that experimenters can only take advantage of the printed-word variant under some circumstances. In particular, our findings suggest that printed words can be used to study questions involving phonological representations but perhaps not to study questions involving visual and semantic representations. If it is the case that the use of printed words selectively taps into phonological processing, then it is even possible that this version of the paradigm may be more sensitive to phonological manipulations than the version using pictures. One might be concerned that the printed-word version still needs further evaluation. This is certainly true: After all, the pattern observed in Experiments 3 and 4 here was different from those found in Experiments 1 and 2, which used the more widely-studied picture version. But this concern has already begun to be addressed: McQueen and Viebahn (in press) have shown, using printed words, very similar patterns of phonological effects to those observed with pictures by Allopenna et al. (1998). There is evidence, however, that orthographic information is accessed during spoken-word recognition (e.g., Jakimik, Cole, & Rudnicky, 1985; Seidenberg & Tanenhaus, 1979; Slowiaczek, Soltano, Wieting, & Bishop, 2003; Ziegler & Ferrand, 1998). One challenge for future research within the written-word version of the visual-world paradigm, therefore, will be to separate orthographic from phonological effects.

Our study used the listening-only variant of the paradigm. The results are unlikely to be due to the particular feature. Dahan and Tanenhaus (2005), for instance, using a task in which participants had to move named objects above or below a geometric shape adjacent to the object using a computer mouse, obtained very similar visual form effects to those found by Huettig and Altmann (2004, in press), using the listening-only task. Similarly, Yee and Sedivy (2006), using a task in which participants had to touch one of the displayed objects on a computer screen, observed similar semantic effects to those obtained by Huettig and Altmann (2005), using the simple listening task.

In addition, we used a variant of the paradigm in which, on experimental trials, the entity mentioned in the spoken sentence was not present in the visual display. But each of the three types of effect found here have also been observed when targets have been present (phonological cohort competitor effects, Allopenna et al., 1998; visual-shape competitor effects, Dahan & Tanenhaus, 2005; semantic competitor effects, Yee & Sedivy, 2006). In addition, studies comparing target-present and target-absent conditions (Huettig & Altmann, 2005) have found similar results across these conditions (other than the tendency for fixations to targets to dominate when targets are present). There thus appears to be nothing artificial about the target-absent situation. Given that there were no instructions to manipulate targets on the display, these target-absent trials were in fact not in any way unusual or puzzling to participants. Furthermore, in an equal number of filler trials, targets were present. Participants thus learned that the displays were just as likely to have targets as not. Our results show that, under these conditions, participants search the displays on all trials for possible matches between linguistic and visual information. It just so happened that on the experimental trials targets were absent. If targets had been present on these trials, target matches may have dominated behavior (cf. Fig. 2 in Huettig & Altmann, 2005, which shows that semantic competitor effects were stronger in a target-absent condition than in a condition where both competitors and targets were present). We feel that the target-absent variant used here maximized the opportunity to observe competitor effects, and their time-course.

A final methodological issue concerns the possibility of task-specific strategies in the visual-world paradigm. Perhaps participants adopt a strategy of explicit name retrieval when they are confronted with an array of four pictures, and hence effects observed with this paradigm may not reflect normal language processing. There are a number of arguments against this view. First, effects of word frequency (Dahan et al., 2001a) and of lexical competitors which are absent from the display (Dahan et al., 2001b) suggest that behavior in the visual-world paradigm is not determined solely by the limited contents of the visual display. The effects found in Experiment 1, with absent targets, support this argument. If participants explicitly prenamed the contents of the displays, then one would not expect the systematic patterns of fixations to competitors that were found. Fixation behavior would more likely be random if it were determined by preactivated names that failed to match the spoken words. Second, several studies have shown how manipulations of fine-grained detail in the speech signal (Dahan et al., 2001b; McMurray et al., 2002; Salverda et al., 2003; Shatzman & McQueen, 2006a, 2006b) modulate eye movement behavior in the visual-world paradigm. Since displays were held constant in these experiments it is impossible to attribute these effects to a name preactivation strategy. Third, demonstrations of fixations to visual-form and semantic features, includ-

ing those in Experiments 1 and 2, are inconsistent with the hypothesis that behavior is driven by preactivation of picture names.

It is instructive, with respect to the issue of strategic name retrieval, to compare the results of Experiments 1 and 2 with those of Dahan and Tanenhaus (2005). In that study, the proportion of looks to visual-shape competitors did not vary as a function of preview (the display appeared either 1000 ms or 300 ms before the acoustic onset of the target word, which was spoken in isolation). This corresponds to what was observed here for the shape competitors (mean competitor-distractor ratios of 0.6 in Experiment 1, and 0.59 in Experiment 2), but not for the phonological competitors, where reducing the preview effectively obliterated the phonological effect. Dahan and Tanenhaus argue that this pattern would be consistent with an account in which picture names are preactivated, and use their data to argue against this account. As we have already argued, we endorse their conclusion that behavior in the visual-world paradigm cannot be explained in terms of the preactivation of a limited "verification set" of picture names. But what then of our finding that amount of preview modulates the proportion of looks to phonological competitors? We suggest, given the three earlier arguments against prenaming strategies, that this finding instead reflects the normal operation of the spoken-word and picture-recognition systems, to which we now turn.

Process and representation in word and picture recognition

Our research suggests that spoken-word recognition entails the evaluation of multiple lexical hypotheses, at several levels of representation. In line with many models of this process, it appears that different candidate words, if they are consistent with the acoustic-phonetic information in the speech signal, are considered in parallel at a phonological level of representation. This explains why cohort phonological competitors, as in the present experiments, attract fixations. It also appears that processing at this level does not have to be completed before other stored knowledge about words, including the visual features of their referents and their semantic attributes, is retrieved. That is, spoken-word recognition is a cascaded rather than a serial process. While this view is anything but controversial (see McQueen et al., 2003, for review), the data from Experiment 1 offer a very clear demonstration of cascaded processing: Looks to phonological competitors tended to precede looks to shape and semantic competitors, but nonetheless looks to all three types of competitor overlapped in time.

Another conclusion from these data is that the storage and/or the retrieval of phonological knowledge is independent from the storage/retrieval of conceptual knowledge. If lexical knowledge were accessed in an all-or-none manner, such that retrieval of a word's phonological form necessarily entailed retrieval of all other knowledge of a word, including visual and semantic features, then there could be no difference in the time-course of looks to the phonological and other types of competitors. Norris, Cutler, McQueen, and Butterfield (2006), based on an analysis of patterns of cross-modal identity priming and cross-modal associative priming, also conclude that access to a word's phonological representation need not entail retrieval of semantic knowledge associated with that word. It is important to note that we are not arguing here for a strict distinction between perceptual and conceptual components of semantic knowledge. Our stored knowledge about the concept *bean*, for example, evidently includes knowledge about physical properties (e.g., visual attributes, such as the shape of a bean) and about non-physical properties (e.g., functional attributes, such as that beans are edible). While fixations to shape competitors preceded those to semantic competitors in Experiment 2, this was not the case in Experiment 1. It would thus be premature to argue that physical and non-physical properties of concepts are functionally fully distinct.

It is also uncontroversial that processing in visual-word recognition is cascaded (e.g., Coltheart et al., 2001; Plaut et al., 1996). Experiments 3 and 4, with printed-word displays, might have provided evidence for continuous flow of information from these displays to phonological representations and conceptual representations, but did not do so. We suggest that these experiments do not indicate that visual-word recognition is serial (i.e., that phonological processing of a printed word must be completed before conceptual knowledge associated with that word can be retrieved). This would go against the evidence from the visual-word-recognition literature, in particular the evidence that conceptual knowledge can be retrieved directly from orthography (Plaut & Shallice, 1993; Strain, Patterson, & Seidenberg, 1995). Instead, it appears that the absence of preferential fixations to visual-shape and semantic competitors in Experiments 3 and 4 is because of the nature of language-mediated visual search when the array consists of printed words, a topic which we turn to later.

Cascaded processing in picture recognition, however, is more controversial. According to the theory of lexical access in speech production of Levelt et al. (1999), a picture's name is retrieved only if it has been selected for production (e.g., in a picture naming study, only if the participant intends to name the picture). According to this theory speech production is a serial, two-stage process. But in other theories of speech production (e.g., the model proposed by Dell, Schwartz, Martin, Saffran, & Gagnon, 1997) there is at least limited cascade of processing from conceptual to phonological levels. The evidence from Experiments 1 and 2, along with that from a number of other studies (e.g., Griffin & Bock, 1998;

Morsella & Miozzo, 2002; Peterson & Savoy, 1998) suggests that there is indeed some non-seriality in the speech production system. When there was ample time to view the display (Experiment 1), it appears that picture processing did advance as far as retrieval of the pictures' names: There were fixations to all three types of competitor. But when there was only 200 ms of preview before the onset of the critical spoken word (Experiment 2), picture processing still involved retrieval of visual and semantic features to a degree sufficient to influence eye movements, but insufficient retrieval of the pictures' names to influence behavior. We suggest that there were no preferential fixations to the phonological competitors under these conditions because, by the time a picture's name could have been retrieved, the evidence in the speech signal had already indicated that that phonological competitor was not a part of the sentence. Note that these findings once again suggest that lexical representations of conceptual and phonological knowledge are at least partially independent: A word's conceptual features can be retrieved without necessary recovery of its phonological features.

The tug of war in language-mediated visual search

The experiments presented here demonstrate remarkably rich interactions between, on the one hand, phonological, visual-form, and conceptual representations activated by spoken words, and, on the other hand, phonological, visual-form, and conceptual representations of entities in the concurrent visual environment. Several previous accounts of the mapping between language and vision in the visual-world paradigm have been proposed. Allopenna et al. (1998) argued that the probability of fixating a particular visual object reflects the acoustic/phonetic goodness of fit between the spoken word and the name of the object. Dahan and Tanenhaus (2005) proposed that the probability of fixating a particular visual object reflects a match between lexical visual features (activated by the spoken word) and a coarse structural representation of the object, associated with its location. Huettig and Altmann (2005; Huettig et al., 2006) argued that fixation probability reflects the overlap between the conceptual information conveyed by individual spoken words and the conceptual knowledge associated with visual objects.

One of the most important contributions of the present research is that all three of these more simple notions of the mapping process between spoken words and visual objects must be revised. In Experiment 1, we found fixations to pictures of phonological competitors (which Experiment 2 showed probably cannot be explained in terms of visual or semantic matches, but see Footnote 1), to pictures of visual-shape competitors (and correlations of that behavior with ratings of shape similarity) and to pictures of semantic competitors. Control of materials rules out the possibility that these latter two effects could be due to phonological matches or, respectively, semantic or visual matches. We thus conclude that language-mediated visual attention involves multiple matches at phonological, visual feature, and semantic levels of processing.

We have characterized this situation in terms of a three-way tug of war involving these three types of knowledge. Interestingly, there appears to be no tug of war with printed-word displays. There, search depended only on phonological matches. We have suggested that this was because phonological information is the most relevant for a search among printed words. Matches in terms of pictorial visual features are certainly not relevant in this situation. Semantic matches, however, could in principle be used. The fact that there are no preferential fixations to the printed forms of the semantic competitors (especially with extended preview as in Experiment 4) thus suggests that the bias towards phonological competitors is a response to the task situation. Participants appear to focus attention on the possibility of phonological matches in the situation where the display consists of orthographic representations of the sound forms of words.

But if attention tends to be focused on phonology with printed-word displays, why is it not focused predominately on shape features with displays of pictures? This may be because of limitations in attentional focus. It may be easier for participants to focus on phonology with printed-word displays than to focus on visual features with picture displays. This may in part be a consequence of the relative complexity of the two types of display. It may be easier for participants to complete a phonologically-based search of an array of four words than to complete a search of four pictures based on visual features alone. In the more complex case, participants may choose to rely on all possible sources of information. Phonological information (the picture names) could therefore be valuable in this search, especially given the availability of phonological information in the speech signal. Similarly, semantic information could be helpful. Further research is required on this issue. The comparison between picture and printed-word displays certainly suggests, however, that language-mediated visual search is determined, in part, by the nature of the information in the visual display.

We conclude that eye movements during language-mediated visual search depend on establishing matches between information extracted from the visual display and from the speech signal. These matches can be made at phonological, visual-feature and semantic levels of processing. Attentional shifts thus appear to be co-determined by the type of information in the display (i.e., pictures or words), the timing of cascaded processing in the word- and picture-recognition systems, and by the temporal unfolding of information in the speech signal. In

the situation where the display contains pictures, the result of these constraints is a tug of war among fixations determined by phonological, shape and semantic matches between the knowledge derived from the display and knowledge derived from the word that is concurrently being heard.

Acknowledgments

This research project was started when the first author was at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. Additional support was provided to F.H. by a grant from FWO, the Fund for Research Flanders (G.0181.05). We thank Marloes van der Goot, Laurence Bruggeman, Marieke van Heugten and Joris Janssen for their assistance with preparing, running and analyzing these experiments. We also thank three anonymous reviewers for helpful comments. Parts of this research were presented at the 13th European Conference on Eye Movements, Bern, Switzerland, August, 2005, and at the 46th Annual Meeting of the Psychonomic Society, Toronto, Canada, November 2005.

Appendix A

Experimental materials

Spoken word	Shape competitor	Semantic competitor	Phonological competitor	Unrelated distractor
boon (bean)	sabel (sword)	sla (lettuce)	boog (bow)	cello (cello)
hoefijzer (horseshoe)	magneet (magnet)	zadel (saddle)	hoed (hat)	filter (filter)
peddel (paddle)	fluit (flute)	zeilboot (sailing boat)	perzik (peach)	bril (glasses)
ballon (balloon)	zon (sun)	pop (doll)	bad (bath)	deur (door)
raket (rocket)	fles (bottle)	vlieger (kite)	ratel (rattle)	emmer (bucket)
arm (arm)	rietje (straw)	nier (kidney)	artisjok (artichoke)	muts (hat)
hart (heart)	voetbal (football)	gebit (teeth)	hamster (hamster)	bloemkool (cauliflower)
ananas (pineapple)	boei (buoy)	pinda (peanut)	agent (policeman)	spijker (nail)
paleis (palace)	kennel (kennel)	koning (king)	paling (eel)	slee (sledge)
bal (ball)	kers (cherry)	shuttle (shuttlecock)	bank (sofa)	hond (dog)
lelie (lily)	kroon (crown)	cactus (cactus)	lepel (spoon)	mossel (mussel)
kerk (church)	iglo (igloo)	graf (grave)	ketting (chain)	pan (pot)
berg (mountain)	servet (napkin)	wolk (cloud)	bel (bell)	kies (tooth)
boor (drill)	pijl (arrow)	ladder (ladder)	boom (tree)	neus (nose)
bord (plate)	wiel (wheel)	karaf (carafe)	bot (bone)	aap (ape)
koffer (suitcase)	schilderij (picture)	tent (tent)	kompas (compass)	mug (mosquito)
tang (pliers)	broek (trousers)	fietspomp (bicycle pump)	tak (twig)	oor (ear)
liniaal (ruler)	kam (comb)	kubus (cube)	libel (dragonfly)	paprika (pepper)
das (tie)	veer (feather)	trui (jumper)	dak (roof)	trommel (drum)
moer (nut)	donut (donut)	hamer (hammer)	moeder (mother)	laars (boot)
dolk (dagger)	kurkentrekker (corkscrew)	kanon (cannon)	dorp (village)	television (TV)
fakkel (torch)	ijsje (ice cream)	bom (bomb)	fabriek (factory)	knoop (button)
schildpad (turtle)	ton (barrel)	haai (shark)	schip (ship)	penseel (paintbrush)
beker (beaker)	klos (bobbin)	vork (fork)	bever (beaver)	paraplu (umbrella)
hek (fence)	rail (railway line)	sleutel (key)	helm (helmet)	tas (bag)
ketel (kettle)	slot (lock)	vijzel (jack)	kegel (cone)	vos (fox)
riem (belt)	slang (snake)	sandaal (sandal)	riet (reed)	asbak (ashtray)
kogel (bullet)	ui (onion)	speer (spear)	konijn (rabbit)	vest (waistcoat)
kano (canoe)	worst (sausage)	fontein (fountain)	kameel (camel)	tuba (tuba)
matras (mattress)	brief (letter)	kruk (stool)	masker (mask)	trompet (trumpet)
toren (tower)	beitel (chisel)	brug (bridge)	tomaat (tomato)	haas (hare)
boek (book)	kaart (playing card)	potlood (pencil)	boeddha (buddha)	zaag (saw)
silo (silo)	kasteel (castle)	tractor (tractor)	sigaar (cigar)	bed (bed)
tol (top)	aardbei (strawberry)	baby (baby)	tobbe (tub)	bus (bus)
maan (moon)	gulden (guilder)	tornado (tornado)	marionet (puppet)	voet (foot)
zwaard (sword)	pincet (tweezers)	pistool (gun)	zwaan (swan)	ster (star)
vijl (file)	zuil (column)	schaar (scissors)	vijver (pond)	kluis (safe)
soldaat (soldier)	robot (robot)	bijl (axe)	sok (sok)	piano (piano)
dokter (doctor)	kabouter (gnome)	spuit (syringe)	dolfijn (dolphin)	mand (basket)
pen (pen)	sigaret (cigarette)	bureau (desk)	pet (cap)	anker (anchor)

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.

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

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

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.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General, 132*, 163–201.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cognitive Psychology, 42*, 317–367.

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

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

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 810–838.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between processing levels in spoken word production. *Journal of Memory and Language, 38*, 313–338.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action* (pp. 1–58). Hove: Psychology Press.

Huettig, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: evidence from head-mounted eye-tracking. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond* (pp. 187–207). New York, NY: Psychology Press.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition, 96*, B23–B32.

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

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica, 121*, 65–80.

Jakimik, K., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language, 24*, 165–178.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*, 211–240.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: a study of picture naming. *Psychological Review, 98*, 122–142.

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

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

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

McDonald, S. A. (2000). Environmental determinants of lexical processing effort. Unpublished doctoral dissertation, University of Edinburgh, Scotland. Retrieved December 10, 2004, from http://www.inf.ed.ac.uk/publications/thesis/online/IP000007.pdf.

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., 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.

McQueen, J. M. & Viebahn, M. (in press). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 555–563.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

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

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

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

Parker, R. E. (1978). Picture-processing during recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 284–293.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological coding during language production: evidence

for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 539–557.

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

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: a case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377–500.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113*, 426–442.

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

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

Shatzman, K. B., & McQueen, J. M. (2006a). The modulation of lexical competition by segment duration. *Psychonomic Bulletin & Review, 13*, 966–971.

Shatzman, K. B., & McQueen, J. M. (2006b). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68*, 1–16.

Slowiaczek, L. M., Soltano, E. G., Wieting, S. J., & Bishop, K. L. (2003). An investigation of phonology and orthography in spoken-word recognition. *Quarterly Journal of Experimental Psychology, 56A*, 233–262.

Strain, E., Patterson, K. E., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1140–1154.

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

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 371–386.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1–14.

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.