

## BRIEF REPORTS

---

# Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition

DELPHINE DAHAN

*University of Pennsylvania, Philadelphia, Pennsylvania*

and

MICHAEL K. TANENHAUS

*University of Rochester, Rochester, New York*

Participants' eye movements to four objects displayed on a computer screen were monitored as the participants clicked on the object named in a spoken instruction. The display contained pictures of the referent (e.g., a snake), a competitor that shared features with the visual representation associated with the referent's concept (e.g., a rope), and two distractor objects (e.g., a couch and an umbrella). As the first sounds of the referent's name were heard, the participants were more likely to fixate the visual competitor than to fixate either of the distractor objects. Moreover, this effect was not modulated by the visual similarity between the referent and competitor pictures, independently estimated in a visual similarity rating task. Because the name of the visual competitor did not overlap with the phonetic input, eye movements reflected word-object matching at the level of lexically activated perceptual features and not merely at the level of preactivated sound forms.

Recent years have seen growing interest in the interface between visual perception, language, and action (e.g., Henderson & Ferreira, 2004; Spivey, Tyler, Eberhard, & Tanenhaus, 2001; Zelinsky & Murphy, 2000). The fact that visual perception can influence and be influenced by concurrent linguistic input has prompted the emergence of a new paradigm for studying spoken-language comprehension. In the *visual-world* paradigm, pioneered by Cooper (1974) and further developed by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), language processing is studied by examining participants' eye gaze to objects in a display as a spoken utterance, referring to one or more objects, is heard (see also Altmann & Kamide, 1999). For example, the instruction "pick up the candle" may be heard concurrently with the presentation of a display containing a candle, candy, pear, and bottle. The probability of fixating each object over time, as lin-

guistic information unfolds, is taken to reflect the listener's developing interpretation of the linguistic input. This assumption relies on people directing their attention—and most often their gaze—to what is being referred to in the utterance, when that referent is visually available. Saccadic eye movements are fast and ballistic; thus, they provide a real-time measure of linguistic processing. Moreover, time course can be assessed without interrupting the speech stream, and the listener's interpretation can be inferred without requiring a metalinguistic decision. Thus, the visual-world paradigm has strengths that complement more traditional methods used to study speech perception.

The relationship between eye movements and spoken-language processing has perhaps been demonstrated most convincingly in studies of lexical processing where fixations reflect the dynamics of lexical activation (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; McMurray, Tanenhaus, & Aslin, 2002). The acoustic/phonetic goodness of fit between the speech signal and the names associated with the displayed pictures is a major factor in predicting observed fixations. For example, the probability of fixating a pictured object over time reflects the overlap between the spoken input and the name of the object (Allopenna et al., 1998). As the first sounds of the name unfold (e.g., *can* from *can-*

---

We thank Manon van Laer and Tau van Dijck for their assistance in preparing and running these experiments. This work was conducted at the Max Planck Institute for Psycholinguistics (Nijmegen, The Netherlands), with the support of the Max Planck Society and NIDCD Grant DC-05071. Part of this article was presented at the 43rd Annual Meeting of the Psychonomic Society in Kansas City. Correspondence should be addressed to D. Dahan, Department of Psychology, University of Pennsylvania, 3401 Walnut St., Philadelphia, PA 19104 (e-mail: dahan@psych.upenn.edu).

dle), the probability of fixating objects with names consistent with the input (e.g., candle and candy) increases, whereas the probability of fixating objects with inconsistent names (e.g., pear and bottle) decreases. Moreover, lexical frequency modulates the probability of fixating pictured objects; fixations to objects with frequent names increase faster than do fixations to objects with less frequent names (Dahan, Magnuson, & Tanenhaus, 2001; Magnuson, Tanenhaus, Aslin, & Dahan, 2003).

Allopenna et al. (1998) formalized a linking hypothesis relating fixations to the underlying cognitive processes. In their formalization, the probability of fixating a visually present object reflects the degree of activation of the object's name (which itself is determined by the goodness of fit between the spoken word and the object's name) relative to the degree of activation of the displayed alternatives. The better the match, the greater the probability of a fixation. The linking hypothesis was used to generate predicted fixation probabilities over time, on the basis of lexical activation predicted by the TRACE model of spoken-word recognition (McClelland & Elman, 1986). The predictions closely matched the behavioral data (Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001).

The Allopenna et al. (1998) linking hypothesis assumes that each object in the display is indexed by its name and spatial location and that the probability of making a saccadic eye movement toward an object depends on the goodness of fit between the spoken word and the name of that picture relative to the alternatives' fit. This is plausible because participants can usually see the full array of objects before they hear the name of the referent object. This preview minimally provides them with a perceptual analysis of the visual array, allowing them to establish a perceptual map of the display by extracting the visual appearance of some or all of the objects, indexed by their spatial location (see Irwin, 1992; Kahneman & Treisman, 1984). Preview might also encourage participants to name some or all the objects, allowing them to index each object by its name and spatial location. Implicit lexical labeling of pictures in tasks that do not explicitly require object naming has been demonstrated by Zelinsky and Murphy (2000). Although picture-naming times are typically greater than 500 msec, partial activation of phonological representations may take substantially less time—especially if one assumes cascaded processing (Humphreys, Riddoch, & Quinlan, 1988). Thus, the probability of fixating a picture upon hearing the referent's name might primarily reflect the match between the unfolding acoustic input and preactivated names of the pictures, which are active in working memory.

However, displayed objects need not be indexed by their names for eye movements to reflect the goodness of fit between the spoken input and the objects' names. The probability of fixating a picture could reflect the match between the picture—more precisely, its structural representation—and visual representations associated with

activated lexical candidates (i.e., those for which the sound form matches the acoustic input). Object locations in the display would be indexed by structural (and perhaps semantic) representations, rather than by phonological representations (i.e., names). The present study asked whether eye movements to displayed objects generated during the recognition of a spoken word are mediated by structural, as opposed to phonological, representations of these objects.

We monitored eye movements to a computer display with four pictures as participants heard the name of one of the pictures. On critical trials, the referent picture (e.g., the picture of a snake) was presented along with a visual competitor that displayed some visual features of the concept associated with the referent's name (e.g., the picture of a rope, which has the elongated and coiled shape associated with the concept of a snake) and two unrelated distractor pictures (e.g., a couch and an umbrella; see Figure 1 for an illustration). Of interest was whether participants, upon hearing the referent's name (e.g., *snake*), would fixate the competitor picture (e.g., the rope) more than either of the distractor pictures. Neither the name of the competitor picture nor that of the distractor pictures matched the referent's name. Thus, observing more fixations to the competitor picture than to either of the distractors could not be explained by a match between the spoken input and a preactivated phonological representation of the competitor picture. However, such a visual competition effect could be explained in terms of a match between the conceptual and visual representations associated with the unfolding spoken word and a coarse structural representation of the competitor picture. This view also predicts that visual competition should increase as the competitor's similar-

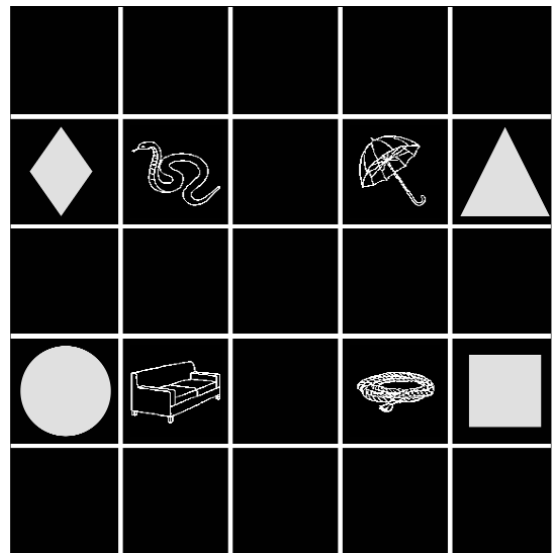


Figure 1. Example of a display.

ity to a prototypical visual representation of the named object increases, above and beyond any effects due to similarity between the referent and competitor pictures.

Which representations (visual or phonological) mediate eye movements to object locations may vary across experimental conditions. In particular, increased viewing time prior to hearing the referent's name might increase the likelihood of implicit naming. In previous studies, preview was about 1 sec (Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Swingley, Tanenhaus, & Magnuson, 2000). To examine whether the amount of pre-exposure affects the representations mediating eye movements, preview time was varied in the present study (300 vs. 1,000 msec). Assuming that activation of the phonological representation associated with a displayed picture takes time to develop and/or requires attention to be directed to the picture location, long preview provides more opportunities for eye movements to be directed to the competitor picture and more time for the development of activation of its phonological representation. Consequently, if eye movements are primarily mediated by phonological representations associated with each picture location, the tendency to fixate the competitor picture more than the distractors should decrease as preview time increases.

## METHOD

### Participants

Forty native speakers of Dutch at the University of Nijmegen participated in this experiment. Half were tested in the 300-msec preview condition and half in the 1,000 msec preview condition.

### Materials

Twenty-two pairs of semantically unrelated picturable concepts were selected. The picture chosen for one of the concepts shared some features associated with a prototypical visual representation of the other. For example, the pair *snake-ropes* was selected because the picture of a coiled rope shares some features with the visual representation most often associated with the concept of a snake. When selecting pictures, we sought to minimize their visual similarity so that the objects could be easily differentiated. For example, we chose a snake in a noncoiled position. Thus, visual similarity was maximized between the visual representation of one of the concepts, the referent, and the picture associated with the other concept, the competitor, and minimized between the competitor picture and the picture chosen for the referent concept. Pictures were black-and-white line drawings, selected from various databases (in particular, Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; Snodgrass & Vanderwart, 1980). The materials are presented in the Appendix.

Similarity between the visual representation associated with one of the concepts and its competitor picture was fairly symmetric for some pairs, but not others. For example, the picture of a mushroom was selected to resemble the visual representation associated with the concept of a lamp, but the selected picture of the lamp did not resemble the visual representation associated with the concept of a mushroom. Thus, upon hearing the spoken word *lamp* and activating the concept of a lamp, participants should fixate the competitor picture "mushroom" more than they should fixate the competitor picture "lamp" upon hearing the spoken word *mushroom*. Both of these conditions were tested. For example, in the condition in which

the visual similarity between the competitor picture and the visual representation associated with the referent's name had been maximized during the construction of the materials (hereafter, the "strong" condition), the picture of the snake was the referent and that of the rope was the competitor. When the roles were reversed (hereafter, the "weak" condition), the picture of the rope was the referent and that of the snake was the competitor. Because some but not all pairs were symmetric, we predicted a competition effect in both conditions, with a smaller effect in the weak condition.

Naming agreement on all  $22 \times 2$  pictures was collected from 15 different participants. Their responses were coded as correct if they corresponded to the intended name. (Small variants of the intended name, such as *handtas* [literally "handbag"] for the intended *tas* [purse, literally "bag"] were treated as equivalent.) Pictures were correctly labeled 89% of the time. Misnaming was largely due to the use of near-synonyms (e.g., the yacht [*yacht*] was also labeled using the Dutch equivalents of boat and ship). Incorrect responses reflecting misidentification of the object (e.g., calling the light bulb a pear) occurred on only 2.9% of the responses.

Each picture-pair was combined with two semantically unrelated pictures with names phonologically dissimilar to those of the referent or competitor pictures. The test stimuli were intermixed with 28 filler sets, consisting of a referent picture and three visually and semantically dissimilar pictures with phonologically unrelated names.

The names of the referents for the test and filler sets were recorded in isolation and in randomized order by a male native speaker of Dutch on DAT-tape in a sound-attenuated room. The recordings were edited using speech-editor software. The mean duration of the spoken words was 620 msec.

**Design and Procedure.** The participants were seated a comfortable distance from the computer screen. Eye movements were monitored with an SMI Eyelink system, sampling at 250 Hz. Spoken stimuli were presented through headphones. On each trial, a central fixation point appeared on the screen for 500 msec, followed by a blank screen for 600 msec. Then, a  $5 \times 5$  grid with four pictures, four geometric shapes, and a central cross appeared on the screen 300 or 1,000 msec before the presentation of the referent's name was initiated (see Figure 1). Prior to the experiment, the participants were instructed to move, on each trial, the named object above or below the geometric shape adjacent to it, using the computer mouse. Positions of the pictures were randomized across four fixed positions of the grid. The positions of the geometric shapes were fixed. The edges of the pictures were approximately 4 cm apart; the distance between the central cross and the closest edge was roughly 3 cm. (One cm corresponded to approximately  $1^\circ$  of visual arc.) The participants were under no time pressure to perform the action. After the participant moved the picture, the experimenter pressed a button to initiate the next trial. Every 5 trials, a central fixation point appeared on the screen, allowing for automatic drift correction.

Condition (strong or weak) was implemented by generating two experimental lists that varied, between items, which of the pair was the trial's referent (e.g., the snake or the rope). Three random orders were created for each list. The same lists were used for both preview conditions. The participants were randomly assigned to preview conditions, lists, and orders.

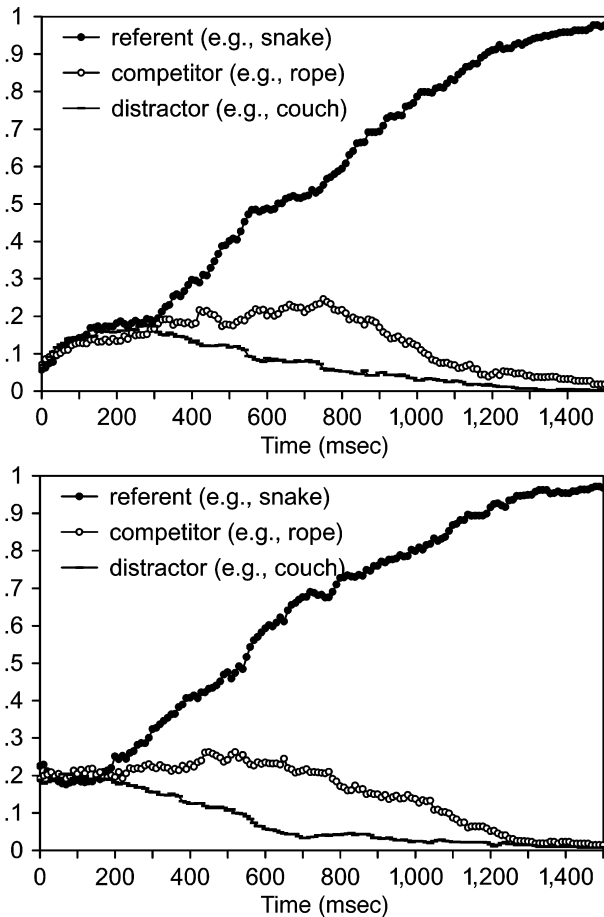
## RESULTS

Data from 9 experimental trials (1% of the data) were excluded because the referent picture was not fixated before or while being moved or because participants moved the wrong picture without correcting their choice.

The data were first parsed into fixations and saccades. Saccade onsets and offsets were automatically detected

using the thresholds for motion ( $0.2^\circ$ ), velocity ( $30^\circ/\text{sec}$ ), and acceleration ( $8,000^\circ/\text{sec}^2$ ). Fixation duration corresponded to the time interval between successive saccades. Fixation location was assessed by averaging the  $x$  and  $y$  coordinates of the fixation's samples and by superimposing the fixation location onto the displayed grid and pictures. Fixations that fell within the grid cell containing a picture were coded as fixations to that picture (see Salverda, Dahan, & McQueen, 2003, for further detail).<sup>1</sup>

We calculated, over successive 10-msec intervals, the fixation proportion to each picture and to the central fixation point for each condition and each participant or item, excluding trials where a saccade or blink occurred during the relevant time interval. Fixation proportions were averaged across participants and across items for subject and item analyses, respectively. Figure 2 presents, for each preview condition, the fixation proportions to each picture type (referent picture, e.g., snake; competitor picture, e.g., rope; and averaged distractor picture, e.g., couch) averaged across participants, from 0



**Figure 2.** Fixation proportions over time from the onset of the referent's name (in msec) to the referent picture, the competitor picture, and the averaged distractor pictures, in the strong condition, with 300-msec preview (upper panel) and 1,000-msec preview (bottom panel).

to 1,500 msec after the onset of the referent picture's name, in the strong condition. Given an estimated 200-msec motor delay for programming and launching an eye movement, the first signal-driven fixations should not begin until 200 msec after the onset of the referent's name (e.g., Hallett, 1986). At around 300 msec in the 300-msec preview, and 200 msec in the 1,000-msec preview, fixation proportions to the referent picture fixations to the competitor picture began to rise and diverge from those to the distractors. Thus, pictures sharing visual similarity with the visual representation associated with the referent's name were fixated more than were distractor pictures.

In order to evaluate this effect statistically, mean fixation proportions to each picture type were computed over a time interval starting from 200 msec and extending over the average duration of the name of the referent (600 msec)—a window from 200 to 800 msec. In order to compare fixation proportion with the competitor and distractor pictures without violating the assumption of independence between samples, we computed the ratio between the proportion of competitor fixation and the sum of competitor- and distractor-fixation proportions, and compared the mean ratio—averaged over participants or items—to .50. A ratio greater than .50 indicates that, out of the fixations devoted to the competitor and either distractor, the competitor received more than half of those fixations. A one-sample  $t$  test revealed that the competitor picture was fixated significantly more than either of the distractor pictures [with a mean ratio of .70,  $t_1(39) = 8.6, p < .001$ ;  $t_2(21) = 5.7, p < .001$ ]. Two-way [Preview (300 vs. 1,000 msec)  $\times$  condition (strong vs. weak)] analyses of variance (ANOVAs) with participants and items as random factors were also conducted on the ratio values, reported as  $F_1$  and  $F_2$ , respectively. The ANOVAs revealed a main effect of condition, with higher ratios in the strong condition than in the weak condition [.73 vs. .68;  $F_1(1, 38) = 6.2, p < .05, MS_e = .0069$ ;  $F_2(1, 20) = 5.5, p < .05, MS_e = .0130$ ] and a main effect of preview, with higher ratios with the 1,000-msec preview than with the 300-msec preview [.82 vs. .59;  $F_1(1, 38) = 53.4, p < .0001, MS_e = .0192$ ;  $F_2(1, 20) = 4.9, p < .05, MS_e = .0238$ ]. The interaction between preview and condition was significant only by participants, with a large difference between the strong and weak conditions in the 300-msec preview (.63 vs. .55), but no such difference in the 1,000-msec preview [.82 vs. .81;  $F_1(1, 38) = 4.9, p < .05, MS_e = .0069$ ;  $F_2 < 1$ ]. One-sample  $t$  tests confirmed that the ratios in the 300-msec and 1,000-msec previews were significantly greater than .50 [ $t_1(19) = 3.4, p < .005$ ;  $t_2(21) = 3.3, p < .005$ ; and  $t_1(19) = 19.9, p < .001$ ;  $t_2(21) = 5.3, p < .005$ , respectively].

These results indicate a clear visual competition effect. Competitor pictures were fixated more than were distractor pictures even though the names of the competitor pictures did not overlap with the phonetic information. The greater difference between competitor and distractor fixations in the strong condition than in the

weak condition was predicted and reflects the procedure used to select the materials. The materials were constructed to maximize similarity between the picture chosen for one of the concepts and the visual representation associated with the other concept, and not vice versa. More important, the visual competition effect was found for both preview conditions, and in fact was larger when the preview was extended. There is thus no evidence that giving participants a longer pre-exposure with the display resulted in an overall decrease of the visual competition effect. Even more striking, a contingency analysis on the 1,000-msec preview condition showed that whether or not the competitor picture had been fixated during preview did *not* affect whether or not the competitor picture was fixated during the rest of the trial [ $\chi^2(1) = 0.044, p > .83$ ].<sup>2</sup> This last analysis suggests that even when participants' attention had been directed to the competitor picture during preview, they did not activate its phonological representation, or at least did not recruit this phonological representation when later searching for the referent of the spoken word.

A possible counterargument to this conclusion is that visual similarity between the competitor and the referent pictures on the display caused participants to misidentify and misname the competitor picture as the referent on some of the trials. Thus, the visual competition effect would originate from the visual similarity between the referent and competitor pictures, rather than, as we argue, the visual similarity between the competitor picture and the visual representation of the concept activated by the referent's name. The difference between the strong and weak conditions speaks against this argument: It is unclear how the visual similarity between displayed objects could be asymmetrical. Furthermore, we directly addressed the contribution of the visual similarity between the referent and competitor pictures on competitor fixations by collecting picture similarity ratings from an independent group of 16 Dutch speakers. Participants saw a series of two-picture displays and gave a visual similarity rating for each picture pair on a scale of 1 (*not similar at all*) to 7 (*very similar*). The correlation between the averaged similarity ratings for the 22 referent-competitor pairs (which varied from 2 to 5.25, with a standard deviation of 0.9) and the size of the visual competition effects (indexed by the ratio between competitor and distractor fixation proportion) was very small and not significant [ $r(20) = .09$ ]. Furthermore, fixation-ratio analyses on low- and high-similarity trials—using a median split across the 22 item pairs according to similarity ratings—showed no significant effect of visual similarity ( $F_1$  and  $F_2 < 1$ ) and a significant visual competition effect on both types of trials. The fixation-proportion ratio was .58 for both high- and low-similarity trials and significantly different from .50 [ $t_1(38) = 3.3, p < .005$ ;  $t_2(10) = 4.0, p < .005$ ; and  $t_1(37) = 2.9, p < .01$ ;  $t_2(10) = 3.8, p < .005$ , respectively].

Taken together, these results demonstrate that, as the name of the referent object was heard, visual features of

the associated concept were activated. Eye movements generated during this activation reflect an ongoing match between activated visual features and a coarse structural representation associated with each of the object locations. The observed visual competition effect is inconsistent with the hypothesis that eye movements merely reflect a match between the unfolding speech and preactivated phonological representations associated with object locations.

## DISCUSSION

The results help clarify the linking hypothesis between lexical processing and fixations to potential referents in the *visual-world* paradigm. Previous studies have shown that, as the name of the referent object unfolds, participants are more likely to fixate an object with a name that matches the phonetic input than an object with an unrelated name. Such fixations are subject to two possible interpretations. Participants could orient their gaze toward an object's spatial location because its structural representation matches the visual representation of the concept activated by the phonetic input. Alternatively, participants could fixate this object's spatial location because its phonological representation (i.e., its name), activated during preview, matches the currently available phonetic input—in effect, bypassing normal lexical processing by using activated phonological representations as a verification set.

Our two main findings strongly support the hypothesis that word-object matching occurs at the level of visual features and not at the level of preactivated sound forms. First, participants fixate visual competitors whose names do not overlap with the name of the referent more than they fixate distractor objects. Second, increasing the amount of preview does not reduce this visual competition effect. Thus, the results contribute to our understanding of what mediates eye movements to displayed objects in the visual-world paradigm, further constraining how to formalize the linking hypothesis between lexical activation and eye movements. Although this conclusion might not hold for experimental conditions that significantly depart from those used here (e.g., more extended preview or repeated exposure of the same pictures), we believe that finding the referent of a linguistic expression in a circumscribed context is similar to that of natural visual search in which the “top-down” target representation, activated from the spoken input, is mapped onto the “bottom-up” scene representation (e.g., Rao, Zelinsky, Hayhoe, & Ballard, 2002).

## 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 & 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.

- COOPER, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84-107.
- CYCOWICZ, Y. M., FRIEDMAN, D., ROTHSTEIN, M., & SNODGRASS, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*, 171-237.
- DAHAN, D., MAGNUSON, J. S., & TANENHAUS, M. K. (2001). 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. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language & Cognitive Processes*, *16*, 507-534.
- DAHAN, D., SWINGLEY, D., TANENHAUS, M. K., & MAGNUSON, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory & Language*, *42*, 465-480.
- HALLETT, P. E. (1986). Eye movements. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10-1, 10-112). New York: Wiley.
- HENDERSON, J., & FERREIRA, F. (2004). *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
- HUMPHREYS, G. W., RIDDOCH, M. J., & QUINLAN, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, *5*, 65-103.
- IRWIN, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 307-317.
- KAHNEMAN, D., & TREISMAN, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. Davies (Eds.), *Varieties of attention* (pp. 29-61). New York: Academic Press.
- MAGNUSON, J. S., TANENHAUS, M. K., ASLIN, R. N., & DAHAN, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*, 202-227.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.
- McMURRAY, B., TANENHAUS, M. K., & ASLIN, R. N. (2002). Gradient effects of within-category phonetic variations on lexical access. *Cognition*, *86*, B33-B42.
- RAO, R. P. N., ZELINSKY, G. J., HAYHOE, M. M., & BALLARD, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*, 1447-1463.
- SALVERDA, A. P., DAHAN, D., & McQUEEN, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51-89.
- SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 174-215.
- SPIVEY, M. J., TYLER, M. J., EBERHARD, K. M., & TANENHAUS, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, *12*, 282-286.
- 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.
- ZELINSKY, G. J., & MURPHY, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, *11*, 125-131.






























#### NOTES

1. The rare fixations that did not fall on one of the pictures or the central fixation point were coded as fixations to the central point.

2. This analysis could not be conducted on the 300-msec preview condition because too few eye movements were performed before word onset (as seen in the low fixation proportions at word onset in Figure 1).

## APPENDIX

Names (with English translations) and pictures of the experimental pairs. The first item corresponds to the referent, and the second corresponds to the competitor in the original condition.

1. trompet (trumpet)	paperclip (paperclip)	12. hoed (hat)	taart (cake)
			
2. ijsje (ice cream)	schroef (screw)	13. schildpad (turtle)	iglo (igloo)
			
3. tas (purse)	slot (lock)	14. fluit (whistle)	plakband (tape)
			
4. slang (snake)	touw (rope)	15. fles (bottle)	kegel (bowling pin)
			
5. lamp (lamp)	paddestoel (mushroom)	16. taart (pie)	arena (stadium)
			
6. riem (belt)	horloge (watch)	17. rups (caterpillar)	trein (train)
			
7. citroen (lemon)	wiel (wheel)	18. wortel (carrot)	fakkel (torch)
			
8. strijkijzer (iron)	jacht (yacht)	19. ui (onion)	tol (top)
			
9. gloeilamp (light bulb)	luchtballon (hot air balloon)	20. bel (bell)	mutts (woolen cap)
			
10. spuit (syringe)	raket (rocket)	21. cocktail (cocktail glass)	trechter (funnel)
			
11. pistool (pistol)	fohn (hair dryer)	22. bom (bomb)	parfum (atomizer)
			

(Manuscript received July 8, 2004;  
revision accepted for publication September 17, 2004.)