Research Article

LINGUISTICALLY MEDIATED VISUAL SEARCH

Michael J. Spivey,[1] Melinda J. Tyler,[1] Kathleen M. Eberhard,[2]
and Michael K. Tanenhaus[3]

[1]Department of Psychology, Cornell University; [2]Department of Psychology, University of Notre Dame; and [3]Department of Brain & Cognitive Sciences, University of Rochester

Abstract—*During an individual's normal interaction with the environment and other humans, visual and linguistic signals often coincide and can be integrated very quickly. This has been clearly demonstrated in recent eyetracking studies showing that visual perception constrains on-line comprehension of spoken language. In a modified visual search task, we found the inverse, that real-time language comprehension can also constrain visual perception. In standard visual search tasks, the number of distractors in the display strongly affects search time for a target defined by a conjunction of features, but not for a target defined by a single feature. However, we found that when a conjunction target was identified by a spoken instruction presented concurrently with the visual display, the incremental processing of spoken language allowed the search process to proceed in a manner considerably less affected by the number of distractors. These results suggest that perceptual systems specialized for language and for vision interact more fluidly than previously thought.*

Prominent theories of visual perception and attention posit that the visual system is functionally independent of other cognitive processes (Pylyshyn, 1999; Zeki, 1993), a modularity thesis that has also been applied to accounts of language processing (Chomsky, 1965; Fodor, 1983). However, a frequently cited counterexample is the fact that visual information regarding a speaker's mouth shape has a powerful influence on speech perception (Massaro, 1997; McGurk & MacDonald, 1976). Moreover, recent eyetracking studies have provided evidence that visual scene perception can constrain real-time comprehension of spoken sentences. For example, temporary ambiguities in spoken word recognition and in syntactic parsing are quickly resolved by information in the visual context (Allopenna, Magnuson, & Tanenhaus, 1998; Spivey, Tanenhaus, Eberhard, & Sedivy, in press; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The present experiments demonstrate the inverse: that language processing can constrain visual perception. In a standard visual search task in which the target object is defined by a conjunction of features, reaction time increases linearly with the number of distractors, often in the range of 15 to 25 ms per item (e.g., Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Wolfe, 1994). However, when we presented the visual display first, and then provided the target features incrementally via a natural spoken query, we found that reaction time was considerably less sensitive to the total number of distractors.

The steep reaction-time-by-set-size function obtained with conjunction search displays was originally interpreted as evidence for serial processing of the objects in a display, and contrasted with the near-flat function relating reaction time to set size in the case of feature search displays—in which a single feature is sufficient to identify the target object. It was argued that the early stages of the visual system process individual features independently and in parallel (Livingstone & Hubel, 1988), allowing the target object to "pop out" in the display if it is discriminable by a single feature, but requiring application of an attentional window to the individual objects, one at a time, if the target object is discriminable only by a conjunction of features (Treisman & Gelade, 1980). This categorical distinction between parallel search of single-feature displays and serial search of conjunction displays has been supported by research using positron emission tomography (PET). PET scans have provided evidence for a region in the superior parietal cortex that is active during conjunction search for motion and color, but not during single-feature search for motion or for color (Corbetta, Shulman, Miezin, & Petersen, 1995).

However, several studies have discovered particular conjunctions of features that do not produce steeply sloped reaction-time-by-set-size functions (McLeod, Driver, & Crisp, 1988; Nakayama & Silverman, 1986; Theeuwes & Kooi, 1994). Additionally, it is possible to observe the phenomenology of pop-out while still obtaining a significant (albeit small) effect of set size on reaction time (Bridgeman & Aiken, 1994). Moreover, it has been argued that steeply sloped reaction time functions may not reflect serial processing of objects in the display, but rather noise in the human visual system (e.g., Eckstein, 1998; Palmer, Verghese, & Pavel, 2000). Overall, a wide range of studies have suggested that the distinction between putatively serial and parallel search functions is continuous rather than discrete, and that such labels should be considered extremes on a continuum of search difficulty (e.g., Duncan & Humphreys, 1989; Nakayama & Joseph, 1998; Olds, Cowan, & Jolicoeur, 2000; Wolfe, 1994, 1998; see also McElree & Carrasco, 1999).

The purpose of the present study was to examine whether incremental processing of natural linguistic input can convert a difficult conjunction search into a pair of easier searches. That is, if a spoken noun phrase such as "the red vertical" is processed incrementally (cf. Altmann & Kamide, 1999; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Marslen-Wilson, 1973, 1975), and there is extremely rapid integration between partial linguistic and visual representations, then the listener should be able to search items with the first-mentioned feature before even hearing the second one. If the observer can immediately attend to the subset of objects sharing that first-mentioned feature, such as the target color (cf. Egeth, Virzi, & Garbart, 1984; Friedman-Hill & Wolfe, 1995; Kaptein, Theeuwes, & van der Heijden, 1995; Motter & Holsapple, 2000), and subsequently search for the target object in that subset upon hearing the second-mentioned feature, then this initial immediate group selection should reduce the effective set size to only those objects in the display that share the first-mentioned feature.

EXPERIMENT 1

Method

In our first experiment, exactly the same visual displays and prerecorded speech cues were used in two subtly different conditions. In

---

Address correspondence to M.J. Spivey, Department of Psychology, Cornell University, Ithaca, NY 14583; e-mail: spivey@cornell.edu.

one block of trials, the participants were informed of the target's identity before the visual display was presented (auditory-first control condition), and in another block of trials, they heard the critical feature words just after onset of the visual display (audiovisual-concurrent, or A/V-concurrent, condition; see Fig. 1).

Participants were instructed to respond to each display as quickly and accurately as possible by pressing the "yes" button if the queried object was present in the display and the "no" button if it was absent. An initial fixation cross preceded the onset of the visual display in order to direct participants' gaze to the central region of the display. Each stimulus bar subtended 2.8° × 0.4° of visual angle, and neighboring bars were separated from one another by an average of 2° of visual angle. Trials with red vertical bars as targets and trials with green vertical bars as targets, as well as red and green horizontal bars as targets, were equally and randomly distributed throughout each session. All participants had normal or corrected-to-normal vision, and all had normal color perception. The same female speaker recorded all speech files, with the identical 1-s preamble recording, "Is there a . . . ," being spliced onto the beginning of each of the four target queries (in which the two adjectives averaged 1.5 s). (In an early pilot study, the objects were presented in a circle so that each object was equidistant from an initial fixation cross. However, participants reported using an overtly serial strategy of selecting a location on the circle and fixating each object in a clockwise fashion until they found the target object. Therefore, the present studies used the more typical scattered visual search displays, as shown in Fig. 1.)

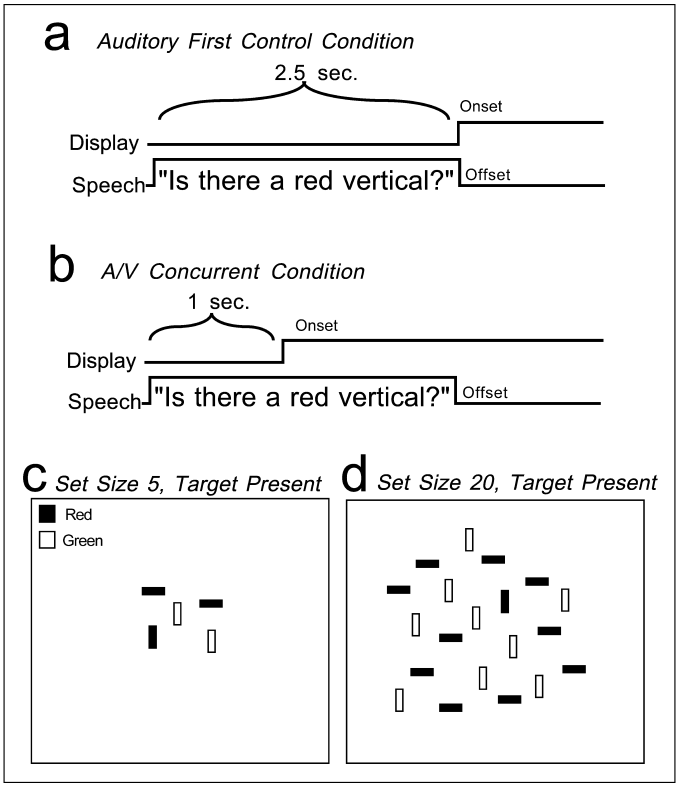

Fig. 1. Examples of the auditory and visual stimuli. In the auditory-first control condition (a), which was similar to standard visual search tasks, the onset of the visual display coincided with the offset of the spoken target query. In the audiovisual-concurrent (A/V-concurrent) condition (b), the onset of the visual display coincided with the onset of the first target-feature word in the spoken query. The example displays show target-present trials with a set size of 5 (c) and 10 (d). In these displays, the target is a vertical red target bar, which is accompanied by vertical green distractor bars and horizontal red distractor bars.

In Experiment 1, each of 20 participants was tested in both an auditory-first control session and an A/V-concurrent session, with order counterbalanced across subjects to compensate for practice effects. The target queries were "Is there a red vertical?" "Is there a green vertical?" "Is there a red horizontal?" and "Is there a green horizontal?" Each session had 96 trials in random order, half with target present and half with target absent; set sizes of 5, 10, 15, and 20 were used.

Results and Discussion

Figure 2 shows the reaction-time-by-set-size functions for target-present trials (filled symbols) and target-absent trials (open symbols) in the A/V-concurrent condition and the auditory-first control condition. Next to each graphed line is the best-fit linear equation, accompanied by its r^2 value, which indicates the percentage of variance

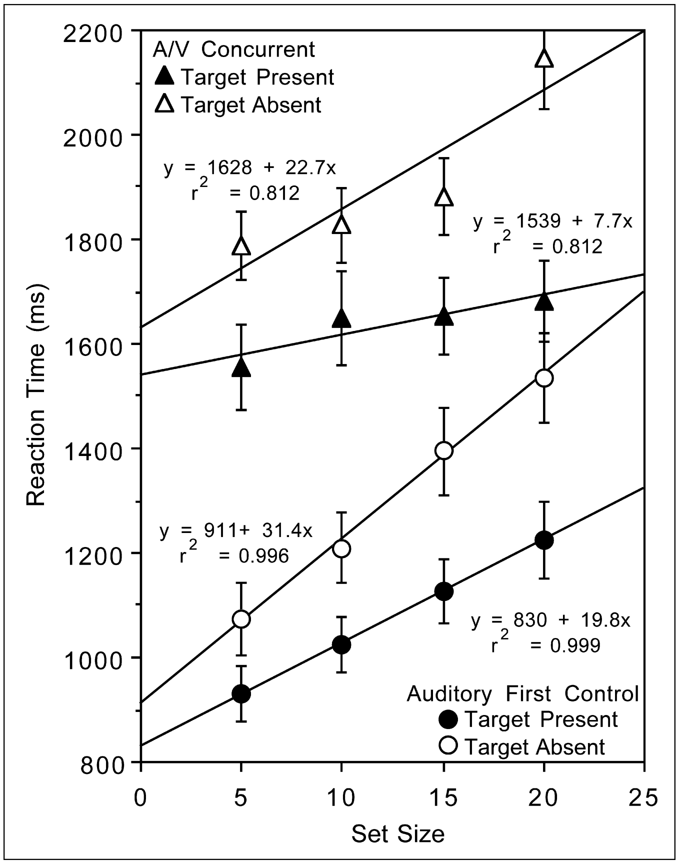

Fig. 2. Results from Experiment 1. The spoken target queries in this experiment had the form "Is there a [color] [orientation]?" Results are shown separately for target-present and target-absent trials for both the auditory-first control condition and the audiovisual-concurrent (A/V-concurrent) condition. Each line is accompanied by the best-fit linear equation and the percentage of variance accounted for (r^2). Error bars indicate ±1 pooled standard error of the mean.

accounted for by the linear regression. In the auditory-first control condition, the reaction-time-by-set-size function was highly linear in both target-present and target-absent trials, as typically observed in standard conjunction search tasks. In contrast, the reaction-time-by-set-size functions for the A/V-concurrent condition were notably less linear. Overall mean reaction time was slower in the A/V-concurrent condition because complete notification of target identity was delayed by approximately 1.5 s relative to the auditory-first control condition. However, because spoken word recognition is highly incremental (Allopenna et al., 1998; Spivey & Marian, 1999), participants were able to begin processing before both target-feature words had been presented, and overall reaction time was delayed by only about 600 ms in the A/V-concurrent condition. Mean accuracy was 96% and did not differ significantly across conditions.

The crucial finding is in the slopes of the reaction-time-by-set-size functions. Despite the fact that the visual displays were identical, results indicated significantly shallower slopes for these functions in the A/V-concurrent condition compared with the auditory-first control condition (see Fig. 2). Repeated measures analyses of variance revealed significant interactions of set size by condition in both target-present trials, $F(3, 57) = 3.51$, $p = .021$, and target-absent trials, $F(3, 57) = 4.51$, $p = .007$, showing that the effect of set size was significantly different in the auditory-first and A/V-concurrent conditions. To specifically test whether the mean slope was shallower in the A/V-concurrent condition, we compared participants' individual set-size slopes from the two conditions via paired *t* tests. These tests revealed significantly shallower slopes for the A/V-concurrent condition than the auditory-first condition in target-present trials, $t(19) = 2.95$, $p = .008$, and even in target-absent trials, $t(19) = 2.21$, $p = .04$.

Although the difference in slopes for target-present trials is striking (7.7 ms/item vs. 19.8 ms/item), the difference in slopes for target-absent trials is more subtle (22.7 ms/item vs. 31.4 ms/item). Moreover, even the slope for the target-present trials of the A/V-concurrent condition was still significantly greater than zero, $t(19) = 3.27$, $p = .004$. In the A/V-concurrent condition, it is quite possible that observers were extracting the appropriate colored bars and, before the second feature was even spoken, immediately locating the odd-one-out in that subset—particularly when the set size was only 5. Such a strategy may have made reaction times with the set size of 5 unusually fast. When the data for set sizes 10 through 20 only were analyzed, the target-present trials in the A/V-concurrent condition exhibited an even shallower mean slope of 3.3 ms/item, which was not significantly greater than zero, $t(19) = 0.75$, $p = .46$.

Overall, the results indicate that simply adjusting the timing of the spoken question (e.g., "Is there a red vertical?") so that the two target-feature words were presented while the visual display was visible allowed participants to find the target object in a manner that was substantially less affected by the total number of distractors. Thus, it appears that in the auditory-first condition, the search process may employ a conjunction template to find the target, thus forcing a serial-like process akin to sequentially comparing each object with the target template. However, in the A/V-concurrent condition, it appears that the incremental nature of the speech input allows the search process to begin when only a single feature of the target identity has been heard. This initial single-feature search proceeds in a more parallel fashion (with the second-mentioned target feature being used to find the target amidst an attended subset), thus dramatically improving the efficiency of search.

EXPERIMENT 2

Method

To further rule out practice and order effects, we replicated the first experiment in a between-subjects design with 10 independent participants in each condition. Additionally, to further test the generality of the results, we reversed the order of the feature words in the spoken query. Each of 20 participants was tested in either the auditory-first control session or the A/V-concurrent session. The target queries were "Is there a vertical red?" "Is there a vertical green?" "Is there a horizontal red?" and "Is there a horizontal green?" This experiment used the same search displays and timing of stimuli as in Experiment 1. Either session had 96 trials in random order, half with target present and half with target absent; set sizes of 5, 10, 15, and 20 were used.

Results and Discussion

Although the spoken queries provided information in the reverse order compared with Experiment 1 (i.e., orientation information came before color information, as in "Is there a vertical red?"), results were essentially the same as before. As expected, the A/V-concurrent condition elicited reaction times that were about 900 ms slower than those in the auditory-first control condition, because of participants having to wait until the orientation adjective was spoken before hearing the onset of the color word. (Note that in Experiment 1, participants had a shorter delay in the A/V-concurrent condition because the first-mentioned target feature was a short, one-syllable color word.) Crucially, the A/V-concurrent condition produced shallower reaction time functions than the auditory-first control condition, particularly in the target-present trials (see Fig. 3). In two-group *t* tests, participants' individual set-size slopes were significantly shallower for the A/V-concurrent condition than for the auditory-first control condition (8.9 ms/item vs. 18.6 ms/item) in target-present trials, $t(18) = 3.55$, $p = .002$, but in target-absent trials, set-size slopes were only marginally shallower for the A/V-concurrent condition than for the auditory-first control condition (21.6 ms/item vs. 30.6 ms/item), $t(18) = 1.9$, $p = .073$. Mean accuracy in this experiment was 94% and did not differ significantly across conditions.

EXPERIMENT 3

Method

In an additional silent control experiment, participants were visually informed of the target's identity at the beginning of each trial. This experiment provides a baseline of performance for these particular displays in a purely visual format, as is typical of the standard visual search paradigm. Ten independent participants were tested in this visual-only control condition. Instead of a speech file informing the participant of the target's identity, an image of the target object (i.e., a red vertical bar, a red horizontal bar, a green vertical bar, or a green horizontal bar) was presented in the center of the screen for 1 s prior to onset of the search display. This experiment used the same search displays as in Experiments 1 and 2. The session had 96 trials in random order, half with target present and half with target absent; set sizes of 5, 10, 15, and 20 were used.

Results and Discussion

This visual-only control condition produced results that were almost identical to those observed in the auditory-first control condi-

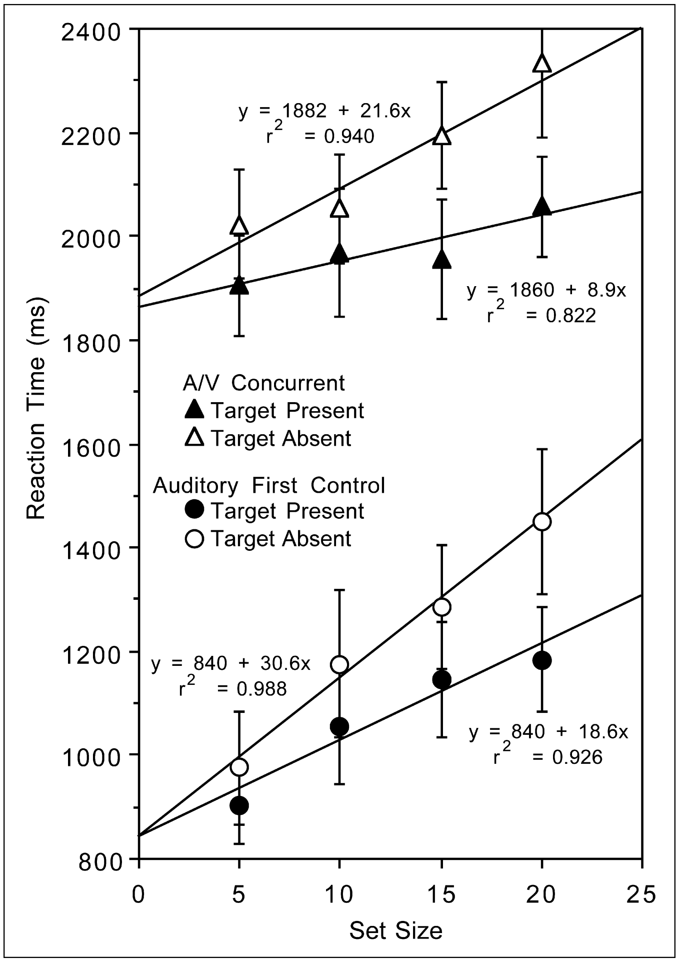

Fig. 3. Results from Experiment 2. The spoken target queries in this experiment had the form "Is there a [orientation] [color]?" Results are shown separately for target-present and target-absent trials for both the auditory-first control condition and the audiovisual-concurrent (A/V-concurrent) condition. Each line is accompanied by the best-fit linear equation and the percentage of variance accounted for (r^2). Error bars indicate ±1 pooled standard error of the mean.

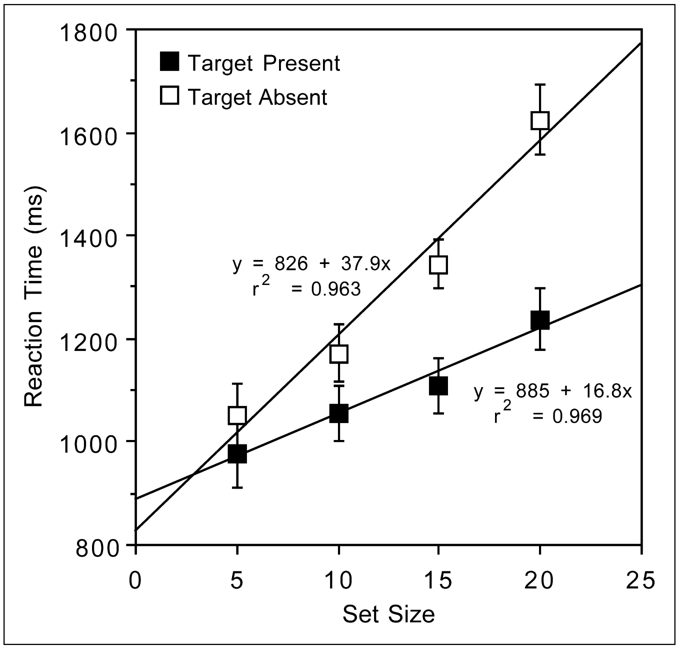

Fig. 4. Results with the visual-only control condition (Experiment 3). Results are shown separately for target-present and target-absent trials. Each line is accompanied by the best-fit linear equation and the percentage of variance accounted for (r^2). Error bars indicate ±1 pooled standard error of the mean.

tions of the previous experiments (see Fig. 4). The reaction time functions from the A/V-concurrent conditions of the previous experiments were significantly shallower than the functions from this visual-only control condition. In two-group *t* tests, the target-present slopes and the target-absent slopes were shallower in the A/V-concurrent condition in Experiment 1 than in this visual-only control experiment, $t(28) = 2.45$, $p = .021$, and $t(28) = 2.36$, $p = .025$, respectively. Similarly, the target-present and target-absent slopes were shallower in the A/V-concurrent condition in Experiment 2 than in this visual-only control experiment, $t(18) = 2.99$, $p = .008$, and $t(18) = 2.61$, $p = .017$, respectively. Mean accuracy in this experiment was 95% and did not differ significantly across conditions.

GENERAL DISCUSSION

It appears that because of the incremental nature of spoken language comprehension (Allopenna et al., 1998; Altmann & Kamide, 1999; Cooper, 1974; Eberhard et al., 1995; Marslen-Wilson, 1973, 1975; Spivey et al., in press; Spivey & Marian, 1999; Tanenhaus et al., 1995), the observer-listeners in the A/V-concurrent condition of Experiments 1 and 2 could selectively attend to the subset of objects that exhibited the target feature that was mentioned first in the speech stream. Upon hearing even just a portion of the second-mentioned target feature a few hundred milliseconds later, the observer-listeners could then locate the conjunction target amidst this attended (spatially noncontiguous) subset. Thus, the incremental nature of the auditorily provided target identity in the A/V-concurrent condition caused the relevant set size to be effectively halved. That is, if the reaction time functions from the A/V-concurrent target-present trials were plotted against the number of objects sharing the first-mentioned feature (half the entire set size, on average), their slopes of 7.7 and 8.9 ms/item would be exactly doubled and thus in the same range as those from the auditory-first and visual-only control conditions (15–20 ms/item).

Previous studies of conjunction search have found that when target identity is held constant throughout an experiment, subjects can essentially ignore distractors that do not exhibit the target color (Egeth et al., 1984; Kaptein et al., 1995). However, this kind of automatic feature selectivity had not yet been shown for orientation until the present study (Experiment 2). Additionally, Olds et al. (2000) have demonstrated that when the conjunction search array itself is presented incrementally, with objects of only one feature type being presented alone for the first 50 ms before full display presentation, reaction times are considerably decreased (see also Watson & Humphreys, 1997, for a study using greater asynchronies in the incremental display presentation). Olds et al. suggested that even though pop-out does not occur in these cases, some partial processing of the target object's identity is carried out during those initial 50 ms and assists the search process overall. Complementing these previous studies, the present findings show that even when the target's identity is unpredictable at the beginning of each trial, and the incrementality of the input comes from natural comprehension of spoken language, observers-listeners can map in real time the incoming linguistic constraints onto the visual array—treating a difficult conjunction search more like a nested pair of easier feature searches. The present results further highlight the incremental processing of spoken language comprehension in general, and specifically demonstrate the human brain's ability to seamlessly cross-index partial linguistic representations (of a noun phrase) with partial visual representations (of a cluttered visual display).

A number of studies have demonstrated facile interaction between functional subsystems of the brain. For example, in contrast to claims of visual modularity (Pylyshyn, 1999; Zeki, 1993), voluntary attention appears to influence early stages of visual processing (Artim & Bridgeman, 1989; Brefczynski & DeYoe, 1999; Gandhi, Heeger, & Boynton, 1998; Motter, 1993; Roelfsema, Lamme, & Spekreijse, 1998; Spivey & Spirn, 2000). Similarly, in contrast to claims of linguistic modularity (Chomsky, 1965; Fodor, 1983), visual context appears to immediately influence spoken word recognition and syntactic parsing (Allopenna et al., 1998; Spivey et al., in press; Tanenhaus et al., 1995). Indeed, much recent work has highlighted the functional interconnectivity between various cortices in vision and in language (e.g., Churchland, Ramachandran, & Sejnowski, 1994; Desimone & Duncan, 1995; Douglas, Koch, Mahowald, Martin, & Suarez, 1995; Pulvermüller, 1999; see also Sekuler, Sekuler, & Lau, 1997, and Vroomen & de Gelder, 2000, for influences of audition on visual perception). The present results show that attentional processes driven by spoken language affect the way people immediately process features in a visual scene, providing further evidence for rapid interaction between linguistic processing and visual perception.

Acknowledgments—We thank Bruce Bridgeman, Shimon Edelman, David Field, Barbara Finlay, Sam Glucksberg, Ulric Neisser, Liddy Olds, Michael Owren, Daniel Richardson, and Julie Sedivy for helpful discussions and comments, and Joshua Richardson and Quinn Hamilton for assistance with pilot studies and data collection. This work was supported by a Sloan Fellowship in Neuroscience (M.J.S.). All participants gave informed consent.

REFERENCES

Allopenna, P.D., Magnuson, J., & 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.

Artim, J., & Bridgeman, B. (1989). The physiology of attention: Participation of cat striate cortex in behavioral choice. *Psychological Research*, *50*, 223–228.

Brefczynski, J.A., & DeYoe, E.A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nature Neuroscience*, *2*, 370–374.

Bridgeman, B., & Aiken, W. (1994). Attentional "popout" and parallel search are separate phenomena. *Investigative Ophthalmology & Visual Science*, *35*, 1623.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Churchland, P.S., Ramachandran, V., & Sejnowski, T. (1994). A critique of pure vision. In C. Koch & J. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 23–60). Cambridge, MA: MIT Press.

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.

Corbetta, M., Shulman, G., Miezin, F., & Petersen, S. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science*, *270*, 802–805.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Douglas, R.J., Koch, C., Mahowald, M., Martin, K., & Suarez, H. (1995). Recurrent excitation in neocortical circuits. *Science*, *269*, 981–985.

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

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

Eckstein, M.P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, *9*, 111–118.

Egeth, H.E., Virzi, R., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 32–39.

Fodor, J.A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Friedman-Hill, S., & Wolfe, J. (1995). Second-order parallel processing: Visual search for the odd item in a subset. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 531–551.

Gandhi, S.P., Heeger, D.J., & Boynton, G.M. (1998). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences, USA*, *96*, 3314–3319.

Kaptein, N.A., Theeuwes, J., & van der Heijden, A.H.C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1053–1069.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*, 740–749.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522–523.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226–228.

Massaro, D.W. (1997). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1517–1539.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for conjunctions of movement in visual search. *Nature*, *332*, 154–155.

Motter, B.C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, *70*, 909–919.

Motter, B.C., & Holsapple, J.W. (2000). Cortical image density determines the probability of target discovery during active search. *Vision Research*, *40*, 1311–1322.

Nakayama, K., & Joseph, J. (1998). Attention, pattern recognition, and pop-out in visual search. In R. Parasuraman (Ed.), *The attentive brain* (pp. 279–298). Cambridge, MA: MIT Press.

Nakayama, K., & Silverman, G.H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*, 264–265.

Olds, E.S., Cowan, W., & Jolicoeur, P. (2000). The time-course of pop-out search. *Vision Research*, *40*, 891–912.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*, 1227–1268.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–336.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case of impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341–423.

Roelfsema, P.R., Lamme, V., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*, 376.

Sekuler, R., Sekuler, A.B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*, 308.

Spivey, M.J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*, 281–284.

Spivey, M.J., & Spirn, M. (2000). Selective visual attention modulates the direct tilt aftereffect. *Perception & Psychophysics*, *62*, 1525–1533.

Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (in press). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology.*

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

Theeuwes, J., & Kooi, K.L. (1994). Parallel search for a conjunction of contrast polarity and shape. *Vision Research*, *34*, 3013–3016.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1583–1590.

Watson, D.G., & Humphreys, G.W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review*, *104*, 90–122.

Wolfe, J.M. (1994). Guided Search 2.0: A revised mode of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.

Wolfe, J.M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*, 33–39.

Zeki, S. (1993). *A vision of the brain*. Oxford, England: Blackwell Scientific.

(RECEIVED 8/30/00; ACCEPTED 10/13/00)