

Linguistically mediated visual search: The critical role of speech rate

BRADLEY S. GIBSON, KATHLEEN M. EBERHARD, and TED A. BRYANT
University of Notre Dame, Notre Dame, Indiana

Recent studies have shown that the presentation of concurrent linguistic context can lead to highly efficient performance in a standard conjunction search task by the induction of an incremental search strategy (Spivey, Tyler, Eberhard, & Tanenhaus, 2001). However, these findings were obtained under anomalously slow speech rate conditions. Accordingly, in the present study, the effects of concurrent linguistic context on visual search performance were compared when speech was recorded at both a normal rate and a slow rate. The findings provided clear evidence that the visual search benefit afforded by concurrent linguistic context was contingent on speech rate, with normal speech producing a smaller benefit. Overall, these findings have important implications for understanding how linguistic and visual processes interact in real time and suggest a disparity in the temporal resolution of speech comprehension and visual search processes.

A number of studies have demonstrated that a concurrent visual context can constrain many aspects of the incremental process of comprehending spoken utterances, including word recognition (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Tanenhaus, & Chambers, 2002; Tanenhaus, Magnuson, Dahan, & Chambers, 2000), syntactic parsing (e.g., Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy, 1995), and pragmatic interpretation (e.g., Barr & Keysar, 2002; Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Sedivy, 2002). Recently, Spivey, Tyler, Eberhard, and Tanenhaus (2001) demonstrated a complementary effect—namely, that the incremental interpretation of spoken utterances can constrain the process of detecting a visual object in a concurrent display. The aim of the present study was to further examine the constraining effects of spoken language on visual perception.

Spivey et al. (2001) employed a visual search task that required participants to detect whether a target defined by a conjunction of color (red or green) and orientation (vertical or horizontal) features was present in displays ranging from 5 to 20 distractors. The critical manipulation involved presenting the search display concurrently with a spoken question that named the target's color and orientation features (e.g., *Is there a red vertical?*). Specifically, the display's onset coincided with the onset of the name of the target's first feature (e.g., *red*). In a control condition, the display's onset occurred immediately after the question (e.g., after *vertical*).

The analyses of the slopes of the linear $RT \times$ display size functions showed that search performance was twice

as efficient in the concurrent condition than in the control condition, regardless of the order in which the target's two features were named. Spivey et al. (2001) attributed the greater efficiency in the concurrent condition to the incremental recognition of the names of the target's features biasing an incremental search of the display. Specifically, the recognition of the name of the first feature (e.g., *red* in *Is there a red vertical?*) caused participants to selectively attend to the subset of items in the display that possessed that feature. This subset, in turn, facilitated the subsequent search for the target's second feature (e.g., *vertical*). Although the efficiency of using a selective-subset search for a conjunction feature target had been demonstrated in previous studies (e.g., Egeth, Virzi, & Garbart, 1984; Friedman-Hill & Wolfe, 1995; Kaptein, Theeuwes, & van der Heijden, 1995), Spivey et al. (2001) were the first to show that the strategy can be induced by the inherently incremental process of recognizing the spoken names of the target's features when the search display is copresent. Conversely, the control condition exhibited less efficient search performance, because there was no bias to use a selective-subset search when the incremental processing of the names of the target's features occurred *before* the search display was presented.

Although the slopes of the linear response time (RT) \times display size functions showed a more efficient search in the concurrent condition, the overall RTs in this condition were significantly longer than those in the control condition. Spivey et al. (2001) attributed the longer overall RTs to the delay in hearing the names of the target's features in the concurrent condition. Specifically, in the control condition, the names of the target's features were heard *before* the onset of the search display, whereas in the concurrent condition, the names were not completely heard until 1.5 sec *after* the onset of the display, which was the average combined duration of the two names. Note that

Correspondence concerning this article should be addressed to B. S. Gibson, Department of Psychology, University of Notre Dame, 118 Haggard Hall, Notre Dame, IN 46556 (e-mail: bgibson@nd.edu).

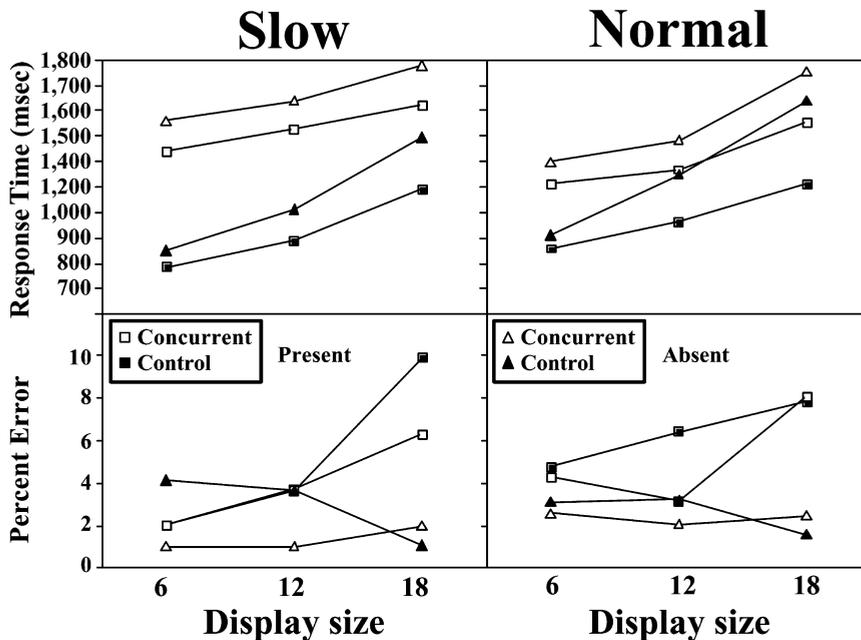


Figure 1. Mean correct response times and percentages of errors in each of the two speech rate conditions (slow vs. normal) depicted as a function of target condition (present vs. absent), question presentation (control vs. concurrent), and display size (6 vs. 12 vs. 18).

the increase in the overall RTs in the concurrent condition was about half the combined duration of the features' names, suggesting that the participants began searching the subset of items for the target's second feature shortly after hearing the initial sounds of the second feature's name. This suggestion is supported by other research showing that spoken words can be recognized as soon as their sounds uniquely identify them, which, in highly circumscribed contexts, can occur before the words' endings (e.g., Allopena et al., 1998; Tanenhaus et al., 1995). Nevertheless, it is important to point out that the average combined duration of 1.5 sec of the names of the target's two features corresponds to a speech rate that is considerably slower than the normal rate. Specifically, the naming of the target's features occurred at an average rate of about three syllables per second, which is about two times slower than the average normal speech rate of five to six syllables per second (Deese, 1984).

The anomalously slow speech rate employed in Spivey et al.'s (2001) study raises the question of whether the efficient search performance in the concurrent condition would occur with a more normal speech rate. A faster, albeit normal, speech rate may result in less efficient search performance if the process of selectively attending to the subset of items possessing the target's first feature is incomplete at the point at which the name of the second feature is recognized. Thus, the potential effect of speech rate was tested by comparing the efficiency of search performance in concurrent and control conditions when the spoken names of the target's two features were produced at a normal speech rate versus the same anomalously slow rate employed in Spivey et al.'s (2001) study.

METHOD

Participants

Twenty-four undergraduates from the University of Notre Dame participated in this experiment to fulfill a course requirement. Twelve participants were assigned to the slow speech rate condition, and 12 were assigned to the normal speech rate condition.

Auditory Stimuli

The auditory stimuli consisted of two sets of digitized recordings created with an artificial speech generation program (MacIntalk) and speech editing software (SoundEdit Pro).¹ Both sets contained recordings of the form *Is there a color orientation?*, where the color word was *red* or *green* and the orientation was *vertical* or *horizontal*, resulting in a total of four color-orientation combinations in each set. The sets differed with respect to the speech rate of the recordings, with one set representing a slow speech rate condition (average of 3.0 syllables per second) and the other set representing the normal speech rate condition (average of 4.8 syllables per second). The durations of the words in the slow rate recordings were comparable to the durations of the words in the recordings used by Spivey et al. (2001). Specifically, the duration of the preamble, *Is there a*, was 1,050 msec, and the average duration of the four combinations of color and orientation words, which were spliced onto the end of the same preamble recording, was 1,452 msec, resulting in a total average duration of 2,502 msec. In the normal speech rate condition, the preamble's duration was 546 msec, and the average duration of the four combinations of color and orientation words that were spliced to the end of the preamble was 1,007 msec, resulting in a total average duration of 1,553 msec. The difference between the average durations of the recordings in the two sets was reliable [paired $t(3) = 14.32, p < .001$, two tailed].

Visual Stimuli

The visual displays were similar to those used by Spivey et al. (2001). They consisted of 6, 12, or 18 colored bars. Each bar subtended 1.24° of visual angle in length and 0.19° of visual angle in

width, and neighboring bars were separated by an average of 2.0° of visual angle. The color of the bars was either red or green, and their orientation was either vertical or horizontal. The visual search displays were divided into target-present and target-absent displays. The target-absent displays always contained equal numbers of either red horizontal bars and green vertical bars or red vertical bars and green horizontal bars. The target-present displays were identical to the target-absent displays, except that one bar (the target) appeared as a unique combination of color and orientation (red horizontal, red vertical, green horizontal, or green vertical). For instance, when the target was either a red vertical bar or a green horizontal bar, the distractors were red horizontal bars and green vertical bars.

Procedure

Each trial began with the presentation of a fixation cross to orient the participant to the center of the display. After 500 msec, the presentation of one of the four recorded search questions began. In the control condition, the fixation cross continued to be displayed until the end of the recorded question, at which point the cross was immediately replaced by the search display. In the concurrent condition, the fixation cross continued to be displayed until the end of the preamble, *Is there a*, at which point the cross was replaced by the search display, which was concurrent with the naming of the target's two features. The control and concurrent conditions were presented in separate blocks of 96 trials, with each block preceded by 18 practice trials. Within each block, the four recorded questions and their corresponding search displays were presented equally often and in a different random order for each participant, thus varying the target of the search from trial to trial. The speech rate of the recorded questions was manipulated between participants, and thus, each participant heard either the slow rate recordings in both blocks or the normal rate recordings in both blocks. For both rates, the order of the concurrent and the control blocks was counterbalanced across participants. In addition, within each block, there were an equal number of target-present and target-absent displays, as well as an equal number of displays with 6, 12, and 18 items.

The participants were instructed to search each display for the target specified in the recorded questions on each trial and to respond by pressing a *yes* button (target present) or a *no* button (target absent). They were encouraged to respond as quickly as possible while keeping errors to a minimum, and their RTs were measured from the onset of the search displays to their pressing a response button.

RESULTS

Mean correct RTs and percentage of error rates in both the slow speech rate and the normal speech rate conditions are shown in Figure 1, plotted as a function of target condition, question condition, and display size. The slopes and intercepts of the RT \times display size functions, as well as associated R^2 values, are listed in Table 1.

We began our analyses by investigating two central questions. First, was search efficiency greater in the concurrent condition than in the control condition, and second, did this benefit depend on the rate at which the names of the target's features were spoken? To address these questions, we performed a $2 \times 2 \times 2 \times 3$ mixed analysis of variance on mean correct RTs and error rates, with speech rate (slow or normal) as a between-subjects factor and question presentation (control or concurrent), target presence (present or absent), and display size (6, 12, or 18) as within-subjects factors.

There was a main effect of question [$F(1,22) = 49.19$, $MS_e = 135,993.58$, $p < .0001$], which indicated that RTs

Table 1
Slopes (in Milliseconds/Item), Intercepts (in Milliseconds), and R^2 Values as a Function of Target Condition and Question Presentation in Each of the Two Speech Rate Conditions in Experiment 1

Question Presentation	Target					
	Target Present			Target Absent		
	Slope	Intercept	R^2	Slope	Intercept	R^2
Slow						
Concurrent	15.00	1,149	1.00	18.00	1,241	0.97
Control	25.42	618	0.97	36.75	614	0.97
Normal						
Concurrent	20.33	968	0.92	29.42	992	0.91
Control	20.25	739	0.99	43.83	641	1.00

were, overall, longer in the concurrent conditions than in the control conditions. As was expected, the speech rate \times question presentation interaction was also significant [$F(1,22) = 4.95$, $MS_e = 55,168.96$, $p < .05$], which indicated that the overall RT difference between the concurrent and the control conditions was larger in the slow speech rate condition than in the normal speech rate condition (531 and 229 msec, respectively, based on the difference between the intercepts of the two instruction conditions). This result provides evidence that the incremental interpretation of the spoken questions guided the participants' search in the concurrent conditions, with the overall search time being affected by the speech rate. Note also that, for both speech rate conditions, the overall average RTs in the concurrent conditions were less than the average combined duration of the names of the target's two features, which was 1,452 msec in the slow rate condition and 1,007 msec in the normal rate condition. Thus, the lower overall RTs suggest that the participants began searching before the offset of the name of the target's second (orientation) feature.

There was also a main effect of display size [$F(2,44) = 95.03$, $MS_e = 25,655.42$, $p < .0001$] and a significant display size \times question presentation interaction [$F(2,44) = 12.27$, $MS_e = 8,599.24$, $p < .0001$], which reflected a larger increase in RTs as a function of display size in the control conditions than in the concurrent conditions. However, a significant speech rate \times question presentation \times display size interaction [$F(2,44) = 3.13$, $MS_e = 8,599.24$, $p < .05$] indicated that the efficiency of the search in the concurrent condition depended on the speech rate with which the questions were presented. As can be seen in both Figure 1 and Table 1, the difference in the efficiency of search performance in the concurrent condition versus the control condition was consistently larger in the slow speech rate condition than in the normal speech rate condition on both target-present and target-absent trials.

Further analyses showed that the effect of display size on RTs was significantly smaller in the concurrent condition than in the control condition when speech rate was slow [$F(2,22) = 21.75$, $MS_e = 2,238.61$, $p < .0001$] but that the effect of display size on RTs did not differ significantly between the two question presentation condi-

tions when the speech rate was normal [$F(2,22) = 2.86$, $MS_e = 6,360.63$, $p > .05$]. Thus, the benefit associated with the concurrent condition appears to depend critically on speech rate. Note, however, that although the interaction between question presentation and display size did not interact with target presence, it is apparent from Table 1 that search performance appeared to be more efficient in the normal concurrent condition (29.42 msec/item) than in the normal control condition (43.83 msec/item) when the target was absent, whereas search appeared to be equally efficient in both the normal concurrent (20.33 msec/item) and the normal control (20.25 msec/item) conditions when the target was present. Separate evaluation of the question presentation \times display size interaction in each of the target-present and the target-absent conditions confirmed this apparent asymmetry [$F < 1$] for target-present trials and [$F(2,22) = 3.62$, $MS_e = 53,910.12$, $p < .05$, for target-absent trials], suggesting that normal concurrent speech did confer some benefit, relative to the normal control condition, but only on target-absent trials. In fact, however, we believe that search performance may have benefited from the concurrent processing of normal rate speech on both target-present and target-absent trials, but only when the subset of items matching the name of the target's first feature was small. Indeed, as is shown in Figure 1 and Table 1, both the target-present and the target-absent RT \times display size functions were less linear in the normal concurrent condition than in any of the other conditions in this experiment.

To assess this possibility, we reevaluated the interaction between speech rate, question presentation, target presence, and display size by including only display sizes 6 and 12 in the analysis and, then, by including only display sizes 12 and 18 in the analysis. When only display sizes 6 and 12 were included, the question presentation \times display size interaction was significant [$F(1,22) = 11.40$, $MS_e = 5,737.33$, $p < .005$], but the speech rate \times question presentation \times display size interaction was not [$F(1,22) = 1.45$, $MS_e = 5,737.33$, $p > .20$]. In addition, none of the other three-way or four-way interactions was significant (all $ps > .10$ or more). These findings suggest that search performance was benefited equally by concurrent presentation of questions presented with either a normal or a slow speech rate when the display size was smaller. In contrast, when only display sizes 12 and 18 were included, the two-way interaction between question presentation and display size was significant [$F(1,22) = 5.08$, $MS_e = 8,074.33$, $p < .05$], and the three-way interaction between speech rate, question presentation, and display size was also significant [$F(1,22) = 6.94$, $MS_e = 8,074.33$, $p < .02$]. None of the other three-way or four-way interactions were significant (all $ps > .30$ or more). Contrary to the results obtained when there were relatively few elements in the visual search display, these findings suggest that search performance was benefited by concurrent presentation of only the questions presented with a slow speech rate.²

With respect to error rates, the analysis showed that the participants committed more errors overall on target-

present trials than on target-absent trials [$F(2,22) = 27.69$, $MS_e = 21.36$, $p < .0001$], more errors overall in the control condition than in the concurrent condition [$F(2,22) = 4.39$, $MS_e = 18.35$, $p < .05$], and increasingly more errors as display size increased [$F(2,44) = 5.03$, $MS_e = 19.48$, $p < .025$]. As is common in visual search tasks (Chun & Wolfe, 1996), there was also evidence that the participants did not search exhaustively before responding *no*, since the miss rate was found to increase as a function of display size on target-present trials without a concomitant increase in false alarms on target-absent trials [$F(2,44) = 8.42$, $MS_e = 25.56$, $p < .001$, for the target \times display size interaction]. However, there was no evidence that the participants used different trial termination strategies across the concurrent and the control conditions, since the pattern of error rates was similar across these two question conditions. As a result, the target \times question presentation \times display size interaction did not approach significance when error rates were considered in this experiment [$F(2,44) = 1.34$, $MS_e = 20.24$, $p > .25$]. There were no other significant effects (all $ps > .20$ or more).

CONCLUSIONS

The present findings provided clear evidence that the increased efficiency of search performance that results from incrementally recognizing the spoken names of the target's features with a concurrent search display depends on speech rate. Although search was consistently found to be more efficient in the concurrent condition than in the control condition when the names of the target's features were spoken at a relatively slow rate, the difference in search efficiency between the concurrent and the control conditions was less consistent, occurring only when the number of display elements was relatively low, when the names of the target's features were spoken at a normal rate. Altogether, the findings suggest that concurrent linguistic processing does not guarantee improved efficiency of search performance. Rather, the greater efficiency appears to consistently occur only when the rate at which the target's features are named is sufficiently slow to allow the selective attending to the subset of items matching the target's first feature to be completed before the search for the item matching the second feature begins. Moreover, these findings suggest that the time to complete the selection of the subset may increase as the number of items in the display increases.

From a theoretical perspective, it is important to bear in mind that the present study was conducted to better understand the constraining effects of linguistic processing on visual processing. Within this framework, it is noteworthy that limited benefits were observed in the concurrent condition when the search questions were presented at a *normal* speech rate, as opposed to a *fast* speech rate, although we would expect that the benefit of concurrent speech would be further diminished as speech rate increased from normal to fast. Thus, the temporal resolution of the processes underlying a selective-subset search appear to fall, at least partially, outside the

average temporal resolution of the processes underlying speech comprehension. This demonstrated temporal disparity therefore places important theoretical constraints on the extent to which linguistic and visual processes may interact.

In addition, this demonstrated temporal disparity also leads one to question more carefully the nature of the effect that concurrent linguistic processing may have on visual search when benefits are observed. One question of particular importance concerns whether the observed differences between the concurrent condition and the control condition actually reflect differences between incremental and nonincremental search strategies. For instance, Kaptein et al. (1995) have argued that some form of a selective-subset search may occur routinely in conjunction search tasks. In the present study, their conclusion implies that a selective-subset search may have occurred routinely in the both the concurrent and the control conditions. But if this is possible, then what may account for the search benefit observed in the concurrent condition? One possibility is that the concurrent linguistic processing elicits a different form of a selective-subset search than in the control condition.

Indeed, recent research suggests that there may be at least two different subset search strategies (Friedman-Hill & Wolfe, 1995). Although both strategies initially involve restricting the search display to a subset of items and then searching for the unique item within this subset, these strategies differ with respect to the amount of initial processing that is required to select the subset, as well as in the way in which search is conducted on this subset. For instance, the search of the subset may occur serially, with attention narrowly focused on individual items in the subset, or the search may occur in parallel, with attention spread across the spatially noncontiguous items in the subset—which Friedman-Hill and Wolfe referred to as a *second-order parallel search*. Accordingly, conjunction targets will tend to be detected more efficiently as a function of display size when second-order parallel search processes are utilized, because the conjunction target will be unique within the context of the second-order feature and this uniqueness can be detected in parallel independently of display size. However, the second-order parallel search is assumed to require more extensive initial processing of the display, because the spatially noncontiguous items need to be grouped in a way that allows them to be simultaneously compared (Friedman-Hill & Wolfe, 1995). Hence, the increase in search efficiency that is associated with second-order parallel processing is counterbalanced by an overall processing cost that is incurred by the necessity of grouping spatially noncontiguous items.

Moreover, Friedman-Hill and Wolfe (1995) proposed that these two subset search strategies can operate in parallel during visual search, although the serial strategy will tend to be initiated sooner than the parallel strategy. As such, visual search performance may reflect a mixture of these two strategies on any given trial. In partic-

ular, search performance will tend to reflect the operation of the serial search strategy when the task can be completed relatively quickly, because the grouping processes underlying the operation of the second-order parallel search strategy have not yet come to completion—in which case, search will tend to be more strongly affected by display size. In contrast, search performance will tend to reflect the operation of the second-order parallel search strategy when the task cannot be completed quickly—in which case, search performance will tend to be less strongly affected by display size.

On the basis of this proposal, the present findings, as well as the findings reported by Spivey et al. (2001), may be interpreted as follows. In the control condition, search performance may have been affected by display size because the rapidly initiated, but less efficient, serial subset search strategy predominated. However, in the slow speech rate version of the concurrent condition, search performance may have been only weakly affected by display size because the serial search strategy was delayed by the incremental processing of the target's features in the spoken question and, therefore, the slower-to-initiate, but more efficient, second-order parallel search strategy predominated. Furthermore, search performance was intermediate between these two extremes in the normal speech rate version of the concurrent condition, producing results that were similar to the slow speech condition when the number of elements to be grouped was relatively small and producing results that were similar to the control condition when the number of elements to be grouped was relatively large.

In short, the main benefit of concurrent linguistic processing may lie more in its ability to delay visual search processes, thus enabling more time-consuming grouping processes to come to completion before visual search is deployed, than in its ability to induce an incremental subset search strategy per se. Future research will be required to determine the precise nature of the hypothetical grouping and segregation processes that are thought to underlie the second-order parallel search strategy. Fortunately, if the present interpretation is correct, the present paradigm may prove to be especially useful in manipulating the operation of these processes.

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.
- BARR, D. J., & KEYSAR, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory & Language*, *46*, 391-418.
- CHAMBERS, C. G., TANENHAUS, M. K., EBERHARD, K. M., FILIP, H., & CARLSON, G. N. (2002). Circumscribing referential domains in real-time language comprehension. *Journal of Memory & Language*, *47*, 30-49.
- CHUN, M. M., & WOLFE, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*, 39-78.
- DAHAN, D., TANENHAUS, M. K., & CHAMBERS, C. G. (2002). Accent

- and reference resolution in spoken-language comprehension. *Journal of Memory & Language*, **47**, 292-314.
- DEESE, J. (1984). *Thought into speech: The psychology of language*. Englewood Cliffs, NJ: Prentice-Hall.
- EGETH, H. E., VIRZI, R., & GARBART, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 32-39.
- FRIEDMAN-HILL, S., & WOLFE, J. M. (1995). Second-order parallel processing: Visual search for the odd item in a subset. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 531-551.
- 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 & Performance*, **21**, 1053-1069.
- SEDIVY, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory & Language*, **46**, 341-370.
- SPIVEY, M. J., TANENHAUS, M. K., EBERHARD, K. M., & SEDIVY, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, **45**, 447-481.
- 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., MAGNUSON, J. S., DAHAN, D., & CHAMBERS, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, **29**, 557-580.
- 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.

NOTES

1. Comparable results were obtained using digitized recordings of natural speech.
2. Thus, we account for the benefits observed in the normal concurrent condition relative to the normal control condition on target-absent trials, but not target-present trials, in the following way. When display size is 12 or less, search is performed in a more parallel manner in the normal concurrent condition, whereas search is performed in a more serial manner in the normal control condition. In contrast, when display size is 12 or more, search is performed in a more serial manner in both the normal concurrent and the normal control conditions. Because the change from serial search to parallel search is more drastic on target-absent trials than on target-present trials, there can be a greater decrease in the overall slope of the normal concurrent condition, relative to the normal control condition, on target-absent trials than on target-present trials.

(Manuscript received August 6, 2003;
revision accepted for publication May 20, 2004.)