Why Does Language Interfere with Vision-Based Tasks?

Amit Almor

University of South Carolina

Abstract. Conversation with a remote person can interfere with performing vision-based tasks. Two experiments tested the role of general executive resources and spatial attentional resources in this interference. Both experiments assessed performance in vision-based tasks as participants engaged in a language task involving a virtual remote speaker. In both experiments, the language task interfered with the vision task more when participants were speaking or planning what to say next than when they were listening. In Experiment 1, speaking or planning what to say next were also associated with higher interference from a visual distractor than listening, indicating that preparing to speak and speaking pose higher executive requirements than listening. In both experiments, localizing the voice of the remote speaker to the front of participants slightly reduced interference in comparison to other directions. This suggests that remote conversation requires spatial attention resources for representing the position of the remote person.

Keywords: language, vision, central executive, spatial attention

With the increased prevalence of communication by phone and other technologies in modern society, more and more activities that have until recently involved face-to-face communication now involve communication between people that do not share the same physical space. Recent research has shown that the mental demands of remote conversation can interfere with the simultaneous performance of tasks such as driving a car, but the reasons for this interference are not well understood (e.g., Strayer & Johnston, 2001). The research reported here explores two possible explanations for this interference. The first explanation is that, in some cases, language processing can compete with other tasks for central executive resources, leading to interference between language tasks and any task that requires these resources. The second explanation is that, in some cases, language processing can compete with vision-based tasks for specific spatial resources. These two explanations are not mutually exclusive, and both are plausible.

Conversation is a turn-taking activity in which different participants take turns to speak and listen (e.g., Clark, 1996). Before speaking, conversation participants typically plan their next contribution to the conversation. During this planning stage, speakers quickly decide on most of what they want to say (the content) as well as some aspects of how to express this message (the form; Meeuwissen, Roelofs, & Levelt, 2006). The content needs to include sufficient information to convey the intended thought in a relevant, informative, and nonredundant manner. The form has to be interpretable to the other conversation participants but not unnecessarily hard to process (Grice, 1975). Choosing the content and form of utterances can require considerable central executive resources (Keysar, Barr, Balin, & Paek, 1998). Similarly, comprehension frequently involves effortful inferences that also require central executive resources (McKoon & Ratcliff, 1992). This effort may be even higher in remote conversation where speakers have to decide on what to say and how to say it without having a clear representation of the physical context of the listener.

When central executive resources are limited, people may allocate a different amount of these resources to language processing during listening and during speaking or planning what to say next. For example, when a person is performing another task while engaged in conversation, he or she may "tune out" while the other conversation participant is speaking. However, when needing to decide what to say next, especially if answering a specific question, the same person may need to allocate more resources to the conversation and fewer resources to the other task. Thus, a possible explanation for why conversation interferes with other tasks is that language processing requires considerable central executive resources, especially when speakers plan what to say next. It is important to note that planning what to say is an incremental process that does not terminate before speakers begin to speak, but rather goes on during speaking (Ferreira & Swets, 2002; Meeuwissen et al., 2006). Consequently, the mental effort involved in planning the next utterance is likely to affect not only the stage of preparing to speak but also the stage of speaking. Thus, the first specific hypothesis tested here is that the interference posed by language on other tasks will be higher during preparing to speak and speaking than during listening. Comparisons between speaking and listening in terms of visual tasks such as driving have been done before (e.g., Kubose et al., 2006; Meeuwissen et al.,

DOI 10.1027/1618-3169.55.4.260

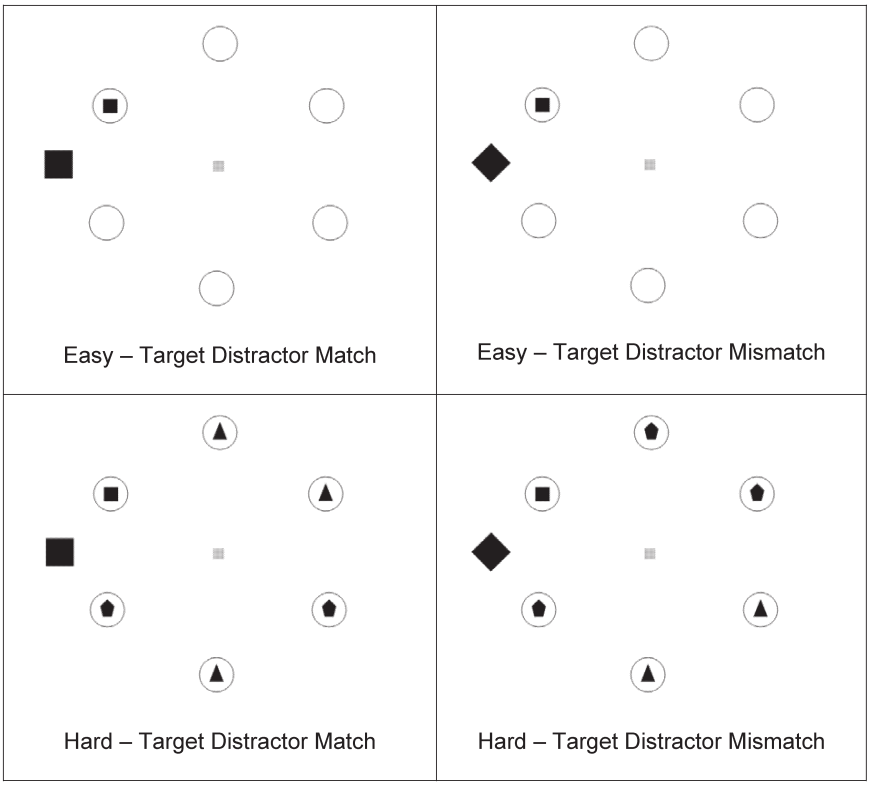

Figure 1. Four displays from the visual target detection task. In all the depicted displays the target shape is the small square. Modeled after Figure 1 in C.S. Green & D. Bavelier, *Nature, 423,* 534 (2003).

2006; Recarte & Nunes, 2003), but not under conditions that resemble the turn taking aspects of normal conversation. The present research used scripted narratives and prompts to engage participants in a language task that mimicked the turn taking of real conversation under carefully controlled experimental conditions.

Conversation can also interfere with vision-based tasks because speakers need to represent and maintain information about what listeners are able to see (Clark, 1996; Clark, Schreuder, & Buttrick, 1983; Horton & Keysar, 1996; Richardson & Dale, 2005; Wilkes-Gibbs & Clark, 1992). For example, a driver who may wish to comment about an exquisite car in the next lane, might say *"that's* what I'd like to be driving" if speaking to a passenger in the car, but would more likely say *"the sports car that just passed me* is what I'd like to be driving" when speaking to someone on the phone. In face-to-face conversation, representing what listeners can see requires speakers to keep track of a listener's position in the shared physical context. Supporting this, Özyürek (2002) has shown that speakers gesture differently when the physical space they share with their listeners is to their front, relative to when it is to their side. Therefore, it is possible that representations of the physical context of other conversation participants involve spatial attentional resources, which can be allocated to specific spatial locations and possibly lead to interference between language and other tasks that require spatial attention resources.

The second specific hypothesis examined here is that keeping track of a listener's position is an obligatory mechanism that operates even in remote communication, such that language processing interferes with a vision-based task *more* when the perceived position of the listener differs from the position to which people need to attend for the visual task than in situations where both positions are the same. This idea is compatible with the view that attention can be directed very much like a spotlight at different aspects of a visual or a mental space (C.W. Eriksen & St. James, 1986; Posner, Snyder, & Davidson, 1980). In partial support of this hypothesis, previous research has shown that the direction from which auditory stimuli are presented in a nonconversational verbal task performed together with operating a driving simulator affects performance in the verbal task, but not in the driving task (Spence & Read, 2003). The present research asked whether the specific demands of linguistic exchange with a remote person can interfere with performing a visual task and whether this interference can be reduced by allowing the attentional spotlight to remain directed at only one area. This was done by virtually positioning the voice of the remote person to coincide with the direction that participants attended for the visual task.

Experiment 1

Experiment 1 tested the effect of language processing on a visual target detection task with a flanker compatibility manipulation (Green & Bavelier, 2003; Lavie & Cox, 1997). This task required participants to indicate which one of two target shapes appeared on a computer screen (see Figure 1). Each display also included a distractor shape that was to be ignored by participants. This distractor sometimes matched the target shape in the trial (matching condition), and sometimes matched the other target shape (nonmatching condition). The number of filled small circles beside the circle containing the target shape made the task easy (no other circles beside the target circle are filled) or hard (all other circles are filled). In visual attention research, the extent to which nonmatching distractors interfere with target detection in comparison to matching distractors is often considered an index of available attentional resources (Lavie, 2005). Attentional resources can be available to process the distractor for one of two reasons. First, it may be that the target detection task does not require much attention because it is perceptually easy, thus leaving sufficient attentional resources free for processing the distractor (Lavie & Cox, 1997). Alternatively, it may be that a secondary task poses high central executive demands that prevent the controlled direction of attention away from the distractor (de Fockert, Rees, Frith, & Lavie, 2001; Lavie, 2005). In summary, when interference is observed in the context of a perceptually easy task, it likely reflects the availability of attentional resources. When interference is observed in the context of performing a hard secondary task, it likely reflects the lack of central executive resources that would be required to direct attention away from the distractor.

The focus of the current research, however, was not on the visual task itself, but rather on how performance of this task was affected by the concurrent language task. This was assessed in two ways. First, the baseline interference imposed by different aspects of language processing on the visual target detection task was gauged by comparing mean response times in the visual task under different conversational conditions. Second, the amount of attentional resources participants directed toward the visual task was assessed by the difference between response times in matching and nonmatching trials. To reduce possible effects of individual differences in conversational styles and to make sure that the language task was not easy and trivial, this study used narratives that were written in advance and were prerecorded. Each narrative was broken into short segments and each segment ended with a question. Although the use of scripted narratives constrained the naturalness of the conversation in this study, it allowed for a high level of uniformity in what different participants heard and in how they were expected to respond. Because previous research has clearly shown that language processing interferes with vision-based tasks (e.g., Strayer & Johnston, 2001), and because the focus of the present research was on differences in the role of specific aspects of language processing in this interference, the present study focused on different language conditions and did not include a no-language condition.

Overall, the predictions for this experiment were as follows. Participants were expected to respond slower in the difficult trials than in the easy trials, slower in the mismatching than in the matching trials, and slower in trials in which they were speaking or preparing to speak than in trials in which they were listening. If speaking and preparing to speak require more central executive resources than listening, participants were expected to: (a) show a larger difference in response time to easy and hard visual trials when they were preparing to speak or speaking than when they listening, and (b) show a larger difference in response times to matching and nonmatching trials when they were preparing to speak or speaking during difficult visual trials than in any other condition. If language processing competes with visual tasks for spatial attention resources, then the visual task should be processed with the greatest ease when both the conversation and the visual task require attending to the same position in space. Therefore, the difference between the matching and nonmatching trials should be the largest when the voice of the remote communicator is heard from the front, which is the same direction participants attended for the visual task.

Methods

Participants

A total of 47 participants (20 male, 27 female) from the University of South Carolina Department of Psychology undergraduate participant pool took part in this experiment after signing an informed consent. Although the age of individual participants was not recorded, all participants were young undergraduate students. The mean age of the psychology departments' participant pool is 20.25 with a standard deviation of .11.

Procedure

Language Task

Four narratives describing how to perform several activities (constructing an igloo, setting up a fish tank, constructing a compost pile, and finding the north using the sun or stars) were digitally recorded and edited into short conversational segments each ending with a question prompt. Each narrative included at least 12 question prompts. The narratives and prompts were designed to engage participants in both listening and speaking but to also vary in level of difficulty. Some of the prompts required recall of information from earlier in the narrative, and some prompts asked for information from the participant's personal experience. This was done to ensure that the results of this experiment could not be attributed to the narratives all being too easy or too hard, or to participants viewing the task as a memory task. Sound editing software was used to create four directional versions of every segment: front (0°), right (90°), rear (180°), or left (270°), which were played during the experiment through a surround sound system. Figure 2 shows the experimental setup. Participants were seated in front of a computer monitor, which presented the stimuli. To discourage participants from identifying the direction from which they heard the voice of the remote person with a loudspeaker, all the loudspeakers were placed behind a screen that prevented participants from seeing them. Participants sat on a chair whose height was individually adjusted to position the participant's head at the level of the speakers. A sound level meter was used to ensure that all versions played at the same volume at the position of participants' head. Each participant heard each of the four scripts from one of the four directions in a counterbalanced fashion, such that each participant heard each script but in a different order and from a different direction. Playing the conversational segments was controlled by an experimenter who was not visible to participants during the experiment and who made sure participants responded appropriately to the conversational prompts. Data from three participants, who the experimenter determined were not engaging the language task, were replaced by data from three new participants. The entire conversation was recorded and later coded for whether the participant was listening, speaking, or preparing to speak. Participants' verbal responses were recorded and analyzed for accuracy. A response was coded as correct if it included only correct information that was relevant to the question/prompt. "Don't know" and "don't remember" responses were coded as wrong.

Target Detection Task

Each trial started with a small gray square appearing at the center of the screen for 250 ms. It was followed by a 500 ms presentation of six empty circles which were then

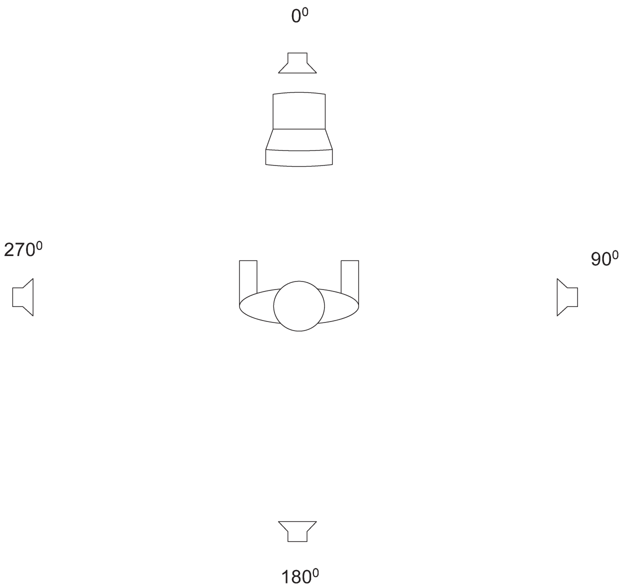

Figure 2. The experimental setup.

replaced with the display of the full stimulus. This stimulus was removed when the participant responded or after a 3000 ms time limit. The target could appear in any one of the six small targets and the distractor always appeared at one side of the display. Only data from trials in which the target and distractor were presented next to one another were analyzed, because only these trials afford clear and straight forward interference (B.A. Eriksen & Eriksen, 1974). The presence of the other trial types ensured that participants could not predict in advance the targets' type and location. Trials of all types were presented in a random order through the testing session. The data from trials in which the participant responded incorrectly, or failed to respond within 3000 ms were removed from further analyses, affecting 8% of the data. An error analysis of performance in the visual task, which is not described here due to space limitations, found no main or interaction effect involving the conversational condition or virtual communicator position factors, thus ensuring that any effects of these factors found in the response time analyses were not due to a speed accuracy trade-off.

Each trial ended with an audio signal that was not audible to the participant but was recorded as a separate channel in the audio recording of the experiment. These periodic signals were later used to synchronize the audio recordings and the target detection data such that the target detection data could be assigned to one of the following two conversational conditions: Listen – a period in which the participant was listening to the prerecorded material; and Speak/Prepare to speak – a period in which the participant was talking as well as the period between the end of the last segment heard by the participant and the onset of the participant's response. Speaking and preparing to speak were not treated as separate categories in this experiment because in this experiment most trials during which participants were preparing to speak also included some speaking. Testing started with a 5-min practice block in which participants performed the visual task without performing the language task.

Due to constraints on the number of observations possible in this task during different conversational conditions, some combinations of conversational condition and virtual speaker position were not represented in the data and, therefore, it was not possible to perform an analysis crossing the effects of these two factors. Thus the effects of these factors were analyzed in two separate analyses, one with *Difficulty*, *Target-Distractor Match*, and *Conversational Condition* as the independent variables, and another with *Difficulty*, *Target-Distractor Match*, and *Virtual Communicator Position* as independent variables. *Target Detection Response Time* was the dependent variable in both analyses.

A combined analysis of conversational conditions and virtual communicator position is addressed in Experiment 2.

Results

Conversational Conditions Analysis

Figure 3 shows response times in the different conditions of the visual target detection task under two conversational conditions (listening vs. preparing to speak or speaking). A repeated measures ANOVA of participants' mean response time (RT) in each condition with factors Difficulty, Target-Distractor Match, and Conversational Condition found a main effect of task difficulty with responses in the hard trials 402 ms slower than responses in the easy trials, $F(1, 45) = 231.27$, $MSe = 64302$, $p < .001$, a main effect of target-distractor match with responses 75 ms slower in the mismatching conditions than in the matching conditions, $F(1, 45) = 11.26$, $MSe = 46110$, $p < .002$, and importantly, a main effect of conversational condition with participants responding 262 ms slower when speaking or preparing to speak than when listening, $F(1, 45) = 144.99$, $MSe = 43424$, $p < .001$. The analysis also found a significant interaction between task difficulty and conversational condition, such that the difference between the easy and hard trials was larger in trials in which participants were speaking or preparing to speak (454 ms) than in trials in which they were listening (350 ms), $F(1, 45) = 7.32$, $MSe = 34027$, $p < .01$. There was a marginally significant interaction between target-distractor match and conversational condition, such that the difference between the matching and nonmatching conditions was larger in trials in which participants were speaking or preparing to speak (121 ms) than in trials in which they were listening (30 ms), $F(1, 45) = 3.44$, $MSe = 46211$, $p < .07$. There was no interaction between task difficulty and target-distractor match, $F = 1.3$, *ns*. The three-way interaction between task difficulty, target-distractor match, and conversational condition was not statistically significant, $F(1, 45) = 2.13$, *ns*. To specifically test the prediction of larger difference between the matching and nonmatching conditions when participants were preparing to speak or speaking during difficult visual trials, a series of planned comparisons contrasted the matching and mismatching conditions for each difficulty and conversational condition combination. These planned comparisons indicated that the difference between the matching and nonmatching conditions was reliable in the difficult prepare to speak/speak condition, $F(1, 45) = 7.34$, $p < .01$, but not in the difficult listen condition, $F < 1$, the easy prepare to speak/speak condition, $F(1, 45) = 1.68$, *ns*, or the easy listen condition, $F(1, 45) = 2.41$, *ns*. This is consistent with the hypothesis that speaking and preparing to speak require more central executive resources than listening. A separate analysis of participants' performance during *listening* segments found no changes in response patterns during these segments. This suggests that the differences between listening and preparing to speak/speaking did not result from a gradual build-up of memory load during listening. The details of this analysis are not described here because of space limitations.

Virtual Communicator Position Analysis

Table 1 shows response times by virtual communicator position. A repeated measures ANOVA with factors Difficulty, Target-Distractor Match, and Virtual Communicator Position confirmed the main effects of task difficulty, $F(1, 44) = 334.87$, $MSe = 79349$, $p < 0.001$, and target-distractor match, $F(1, 44) = 6.87$, $MSe = 38405$, $p < 0.02$, but found no main effect of virtual communicator position, or any interaction effects, all F values < 1 except for the interaction between difficulty and virtual communicator position, $F(3, 132) = 1.85$, *ns*. Although there was no interaction between orientation and target-distractor match, it may be that the predicted increased difference between the matching and nonmatching conditions in the 0° was simply not strong enough to elicit this interaction. To specifically test the prediction that target-distractor match would have the most pronounced effect in the easy 0° condition, a series of planned comparisons contrasted the matching and mismatching conditions for each difficulty and virtual communicator position combination. These planned comparisons revealed a reliable difference between the matching and nonmatching conditions in the easy 0° condition, $F(1, 44)$

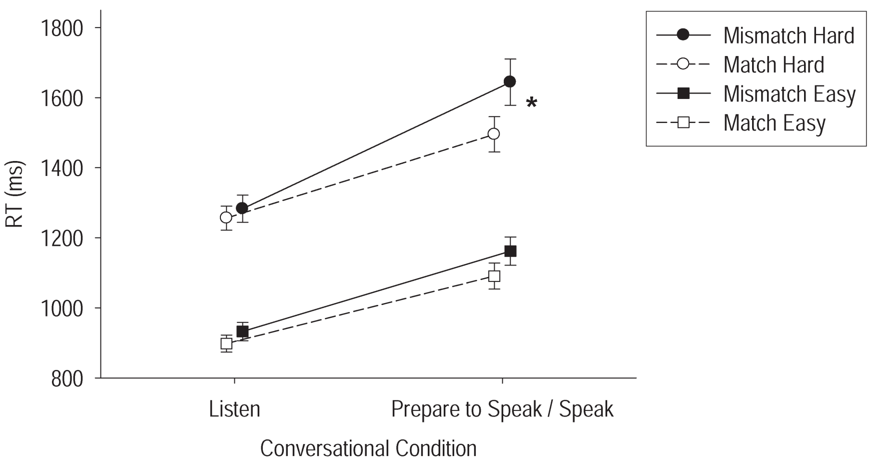

Figure 3. Experiment 1 results shown by conversational condition. Error bars indicate standard error of the means.

Table 1. Mean response times (ms) of Experiment 1 shown by virtual communicator position. Standard errors of the mean are in parentheses

		Virtual communicator position			
Difficulty	Target-distractor match	0°	90°	180°	270°
Easy	Match	918 (24)	933 (25)	933 (27)	944 (35)
	Mismatch	995 (37)	989 (34)	978 (34)	953 (31)
	Matching effect	77*	56	45	11
Hard	Match	1315 (47)	1296 (45)	1383 (43)	1304 (42)
	Mismatch	1325 (43)	1336 (42)	1403 (50)	1354 (53)
	Matching effect	10	40	20	50

*A significant matching effect ($p < .05$) indicated by planned comparisons.

= 4.35, $p < .05$, a marginally reliable difference in the easy 90°, $F(1, 44) = 3.04$, $p < .09$, but no difference between the matching and nonmatching conditions in the easy 180°, $F(1, 44) = 1.72$, *ns*, easy 270°, $F < 1$, or any of the hard conditions, F values < 1.05. The fact that the interference effect in the 0° condition was small made it impossible to test whether it was driven by facilitation in the matching condition or increased interference in nonmatching condition.

An analysis of verbal response accuracy, which is not described here due to space limitations, found no differences in verbal response accuracy between the different virtual communicator position conditions.

Discussion

In addition to the expected difference between the easy and hard conditions, and between the matching and nonmatching conditions, this experiment found strong effects of conversational conditions and a weaker indication of an effect of virtual communicator position. First, with respect to the effects of the language task on mean response times, participants were slower to respond to the visual target task when preparing to speak or speaking than when listening. Participants also showed greater sensitivity to the difficulty of the visual task when speaking or preparing to speak than when listening. This is consistent with the hypothesis that speaking and preparing to speak pose higher central executive demands than listening.

Second, participants only showed a reliable distractor interference effect when speaking or preparing to speak during hard visual trials. This likely indicates that participants were sensitive to the combined difficulty of the language and the target detection tasks, such that when both tasks were difficult (speaking or preparing to speak during hard target detection trials), participants were unable to direct attention away from the distractors. When analyzed by virtual communicator position, an interference effect was found reliably only in the planned comparison for the 0° virtual communicator position during easy trials. However, because the individual comparisons also indicated a marginally significant interference effect in the 90° condition, this result should be considered with caution; any suggestion that spatial attentional demands may play a role in remote conversation should be examined further.

Although these results support the role of central executive resources and, to a lesser extent, spatial attentional resources in the interference between language and vision-based tasks, they do not speak to the question of whether, and how, conversational conditions and virtual speaker position interact. Also, the fact that some of the findings were detected in the planned comparisons, but not in the ANOVAs, raises some concern about the robustness of these findings. Both issues are addressed in the next experiment.

Experiment 2

The second experiment used a pursuit tracking task which is a vision guided action task wherein participants continuously monitor and track a moving target presented on a computer screen. Performance in this task was measured by how closely participants were able to track the moving target. This task assessed the effects of language processing on people's ability to act on the basis of changing visual information without requiring any deliberate decision making. Moreover, because performance in this task could be measured continuously, the data from this task had high temporal resolution. This allowed a fine grain analysis of conversational conditions with speaking and preparing to speak treated as separate conditions. The high number of measurements generated by the high temporal resolution also yielded sufficient data for an analysis of the interaction between conversational conditions and virtual speaker position.

Two main predictions were of primary interest for this experiment. First, if speaking and preparing to speak require more central executive resources than listening, then speaking and preparing to speak should cause greater interference with tracking performance than with listening. Second, if language processing competes with visual tasks for spatial attention resources, then the tracking task should be processed with the greatest ease when both the conversation and the visual task require attending to the same po-

Table 2. Results of Experiment 2 before the log transformation in milliseconds to target. Standard errors of the mean are in parentheses

	Virtual communicator position			
	0°	90°	180°	270°
Listening	2.63 (0.12)	2.60 (0.11)	2.61 (0.13)	2.68 (0.14)
Speaking	2.76 (0.15)	2.71 (0.14)	2.76 (0.14)	2.79 (0.15)
Preparing to speak	2.62* (0.12)	2.95 (0.22)	2.89 (0.19)	3.03 (0.21)

*Significantly shorter time lag ($p < .05$) than in other orientations indicated by planned comparisons.

sition in space. Therefore, the interference from speaking and preparing to speak was expected to be minimal when the voice of the remote conversation participants is heard from the front, which is the same direction participants attended for the target tracking task.

Method

Participants

A total of 47 participants (18 male, 29 female) from the University of South Carolina Department of Psychology undergraduate participant pool took part in this experiment after signing an informed consent.

Procedure

The target pursuit task was controlled by a program which presented a small ball moving smoothly across the screen. To ensure that participants had to continuously monitor the target, the target's motion speed and direction were varied periodically. Participants were instructed to use the computer mouse to keep the cursor as close as possible to the target. The program recorded "time to target" every 10 ms by dividing the distance between the target and the participant controlled cursor by the target's moving speed. The program also emitted a periodic audio signal that was not audible to participants, and was used to synchronize the audio recordings and the tracking data such that the tracking data could be assigned to one of the following conversational conditions: Listen – a period in which the participant was listening to the prerecorded material; Speak – a period in which the participant was talking, and Prepare to speak – the period between the end of the last segment the participant heard and the onset of the participant's response. The conversational task was the same as Experiment 1. Due to the lack of differences between verbal response accuracy in the different virtual communicator positions in Experiment 1, verbal response accuracy was not analyzed in Experiment 2. The independent variables in this experiment were *Conversational Condition* and *Virtual Communicator Position* and the dependent variable was the log transformed (to correct for negative skew) *time to target.*

Results Table 2 shows tracking performance in milliseconds before the log transformation during listening, speaking, and preparing to speak as a function of the position of virtual speaker position. The table shows nontransformed tracking performance because they are easier to interpret than the log transformed times. A repeated measures ANOVA on log transformed time to target measures with factors Conversational Condition and Virtual Communicator Position found a main effect of conversational condition, $F(2, 92) = 13.30$, $MSe = .014$, $p < .001$, no main effect of virtual communicator position, $F < 1$, and a significant interaction of virtual communicator position and conversational condition, $F(6, 276) = 2.24$, $MSe = .012$, $p < .04$. A series of planned comparisons aiming to test the prediction that preparing to speak and speaking pose greater interference than listening further showed that the main effect of conversational condition was driven by pair-wise differences between all the conversational conditions – participants tracked the target more poorly when preparing to speak than when speaking, $F(1, 46) = 5.26$, $p < .03$, and when speaking than when listening, $F(1, 46) = 10.41$, $p < .003$.

Pair-wise planned comparisons were conducted to test the prediction that the interference was minimal when the voice of the remote conversation participants was heard from the front. They showed that preparing to speak interfered with target pursuit tracking more than listening in the 90°, 180°, and 270° virtual communicator positions, $F(1, 46) = 9.95$, $p < .003$, $F(1, 46) = 5.56$, $p < .03$, $F(1, 46) = 12.67$, $p < .001$, but found no difference in the effect of preparing to speak and listening on tracking performance in the 0° virtual communicator position, $F < 1$. There were no significant differences between speaking and the other two conversational conditions in any orientation, F values < 1, likely reflecting the lower power of the orientation-specific comparisons involving the intermediate "speaking" condition.

In this experiment too, there were no changes in performance during *listening* segments, indicating that the differences between the conversational conditions did not result from a gradual buildup of memory load during listening.

Discussion

The results from this experiment replicated and extended the results of the first experiment in that speaking and preparing to speak produced more interference than listening. However, here, preparing to speak produced overall more interference than speaking. The interference from preparing to speak on target tracking was however not present in the 0° virtual communicator position. Thus, positioning the virtual communicator to the front of participants reduced the interference between preparing to speak and performance in a task that did not require explicit decision making. The fact that this experiment produced a similar but more robust effect of virtual communicator position than Experiment 1 lends more credence to the conclusion that, when speakers plan what to say next, they use spatial attention resources in a way that is sensitive to the perceived position of the other conversation participant.

General Discussion

The results from the two experiments clearly show that the mental effort associated with specific aspects of language processing can affect the performance of vision-based tasks. In particular, performance of vision-based tasks appears to be most vulnerable to interference from language when speakers plan what to say next in response to a prompt. This supports the first hypothesis tested in this paper, which was that, during planning, speakers utilize excessive central executive resources.

The second hypothesis tested in this paper was that spatial attention resources are used in conversation with a remote person and that this may lead to interference with vision-based tasks if speakers need to maintain spatially distinct representations for the vision-based task and for the remote person. Compatible with this hypothesis, both experiments found that the interference was reduced when the voice of the remote person was virtually placed in front of participants, thus allowing them to attend to only one "mental space" for both the conversation and the visual task. In Experiment 1, this facilitative effect manifested, albeit weakly, in the perceptually easy conditions. In Experiment 2, this facilitative effect was observed reliably in the hardest conversational conditions. Although these were small effects, their consistency across the two experiments suggests that virtual communicator position can modulate the interference from language processing on vision-based tasks. Together, the two experiments suggest that this interference is high when speakers plan what to say next, but possibly only if they do not need to make a decision or if the vision task is not too hard.

It should be noted that the present results could also be taken to indicate that certain aspects of conversation are easier when the conversational partner's voice is perceived from the front, regardless of where participants have to attend for the visual task. This may reflect people's experience with face to face conversation, in which they are usually able to see the person they are talking to (Özyürek, 2002). The present research shows that, interestingly, these effects occur in remote conversation, in which tracking the gaze of the other conversant is impossible. Future research will have to uncover the specific details of the underlying mechanisms.

Overall, the present results have broad theoretical implications in that they show that language specific processes interact with visual processes and likely use spatial attention resources. Although previous research (e.g., Spence & Read, 2003) made a similar argument about common visual and auditory attentional mechanisms, the differences between the conversational conditions in the present study show that language poses specific attentional demands beyond those associated with processing general auditory information.

Acknowledgments

Thanks to Kristen Simensen Setzler for recording the conversational stimuli and for overseeing other aspects of stimuli preparation. Also thanks to Tim Boiteau, Alison Gulick, Melissa Jantz, Randy Lowell, Jeremy May, Lauren Meara, Melissa Robinson, and Brittany Sheppe for help with stimuli preparation, testing, and transcribing the data. Jeffery Schatz provided the visual attention experiment script. Gordon Baylis, Veena Nair, John Richards, and Brent Strickland made useful comments about earlier versions of this manuscript.

This work was partially supported by a grant from the University of South Carolina Research and Productive Scholarship Fund.

References

Clark, H.H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H.H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning & Verbal Behavior, 22*, 245–258.

de Fockert, J.W., Rees, G., Frith, C.D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science, 291*(5509), 1803–1806.

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*, 143–149.

Eriksen, C.W., & St. James, J.D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics, 40*, 225–240.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*, 57–84.

Green, C.S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534–537.

Grice, H.P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York.: Academic Press.

Horton, W.S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91–117.
Keysar, B., Barr, D.J., Balin, J.A., & Paek, T.S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory & Language, 39*, 1–20.
Kubose, T.T., Bock, K., Dell, G.S., Garnsey, S.M., Kramer, A.F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology, 20*, 43–63.
Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9*, 75–82
Lavie, N., & Cox, S. (1997). On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection. *Psychological Science, 8*, 395–398.
McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440–466.
Meeuwissen, M., Roelofs, A., & Levelt, W.J.M. (2006). Incrementality in naming and reading complex numerals: Evidence from eye tracking. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology, 59*, 296–311.
Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language, 46*, 688–704.
Posner, M.I., Snyder, C.R., & Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160–174.
Recarte, M.A., & Nunes, L.M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied, 9*, 119–137.
Richardson, D.C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*, 1045–1060.
Spence, C., & Read, L. (2003). Speech shadowing while driving: On the difficulty of splitting attention between eye and ear. *Psychological Science, 14*, 251–256.
Strayer, D.L., & Johnston, W.A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*, 462–466.
Wilkes-Gibbs, D., & Clark, H.H. (1992). Coordinating beliefs in conversation. *Journal of Memory & Language, 31*, 183–194.

Received May 8, 2007
Revision received June 9, 2007
Accepted June 15, 2007

Amit Almor

University of South Carolina
Psychology Department
1512 Pendleton Street
Columbia, SC 29208
USA
Tel. +1 803 777-4302
Fax +1 803 777-9558
E-mail almor@sc.edu

