

## **THE EFFECT OF VOICE INTERACTIONS ON DRIVERS' GUIDANCE OF ATTENTION**

Yi-Ching Lee,<sup>1</sup> John D. Lee,<sup>2</sup> and Linda Ng Boyle<sup>2</sup>  
<sup>1</sup> University of Illinois at Urbana-Champaign, <sup>2</sup> University of Iowa  
Human Factors Division  
Institute of Aviation  
One Airport Road  
Savoy, IL 61874 USA  
E-mail: chinglee@uiuc.edu

**Summary:** The objective of the current study was to assess the effect of voice interactions with an in-vehicle system on drivers' guidance of attention. Our approach was to examine the effect of voice interactions on endogenous control of attention using a modified Posner cue-target paradigm. Consistent with the bottleneck hypothesis, dual-task slowing was observed when drivers responded to an auditory task and to a pedestrian detection task concurrently. This interference contributed to disrupted attention allocation, especially when drivers could not rely on their endogenous control of attention.

### **INTRODUCTION**

Motor vehicle crashes are the leading cause of death among Americans 5-44 years old (National Center for Injury Prevention and Control, 2004). In 2004, there were more than 38,000 fatal crashes, resulting in 33,134 deaths, 2,594,000 injuries (US Department of Transportation, 2004), and economic loss of more than \$200 billion (Blincoe et al., 2002). The recent 100-car Naturalistic Driving study concluded that driver inattention is the leading factor in most crashes and near-crashes, where inattentive drivers have a three-times higher risk of near crash/crash risk than attentive drivers (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006).

The growing popularity of in-vehicle information systems could lead drivers to be more inattentive and may undermine safety. For example, the US DOT - National Highway Traffic Safety Administration estimated that 10% of vehicles driven during daylight hours were by someone conversing on a wireless phone (Glassbrenner, 2005). Using wireless devices has the potential to impair driving performance (Alm & Nilsson, 1994) by diverting drivers' attention away from their primary driving task and decreasing drivers' sensitivity to roadway objects (Strayer, Drews, & Johnston, 2003). Speech-based interactions with in-vehicle computers enable drivers to keep their eyes on the road but can still load drivers cognitively (Lee, Caven, Haake, & Brown, 2001). Drivers showed slower braking responses (Levy, Pashler, & Boer, 2006) and slower reaction times to traffic signals (Strayer & Johnston, 2001) and flashing lights (Recarte & Nunes, 2003) when responding vocally to an auditory task than when listening to messages. These results are in line with the predictions from the bottleneck hypothesis (Levy, Pashler, & Boer, 2006; Pashler & Johnston, 1998) such that dual-task slowing is observed in the performance of two concurrent tasks, each requiring a choice of response.

The objective of the current study was to assess the effect of voice interactions with an in-vehicle system on drivers' guidance of attention. Our approach was to examine the effect of voice

interactions on endogenous control of attention using a modified Posner cue-target paradigm. This paradigm is used to assess whether the elevated cognitive load undermines attention allocation, information consolidation, or a combination of both.

## **METHOD**

### **Participants**

Sixteen native English speakers (aged 21 to 30 years) participated in the experiment. They had normal or corrected-to-normal vision, passed Ishihara's tests for color-blindness (Ishihara, 1966), drove at least three times per week and at least 3000 miles per year, and had possessed a valid driver's license for at least 5 years. Participants were compensated for their time. The mean compensation participants received was \$37.25 ( $SD= 1.73$ ).

### **Apparatus and Tasks**

A fixed-based, medium-fidelity driving simulator was used for the experiment. The simulator uses a 1992 Mercury Sable vehicle cab that has been modified to include a screen with a 50-degree visual field of view, force feedback steering wheel, and a rich audio environment. The fully textured graphics are generated by DriveSafety's Vection™ software that delivers a 60-Hz frame rate at 1024 x 768 resolution. Data were collected at a rate of 60 Hz.

Eye movement data were collected at 60 Hz using a Seeing Machines' FaceLab™ eye tracking system (version 4.2). The eye tracking system uses two small video cameras to track head and eye movements and then calculates, among other measures, coordinates for a gaze vector that intersects the simulator screen. The system does not require any head-mounted hardware and is unobtrusive.

The modified cue-target paradigm that included pedestrian crossing signs, pedestrians, and trucks was implemented in simulated driving scenarios. Participants were asked to recognize the location of pedestrian crossing signs and use them to guide their search for pedestrians located in the parking lanes. The specific instructions made the pedestrian crossing signs endogenous cues for detecting pedestrians. Pedestrians were occluded by trucks in the parking lanes and by fog for all but approximately two seconds. When drivers detected pedestrians, they responded by pressing one of the two buttons on the steering wheel.

Participants were in the center lane of a three-lane one-way road with traffic in adjacent lanes. They were asked to use cruise control and follow a lead vehicle traveling at 48.3 kph (30 mph). The lead vehicle braked periodically, and participants needed to brake accordingly and resume their speed and reengage cruise control after each braking event. Additionally, in half of the drives, billboards with high contrast images that flashed 4 times/second were placed on the grass areas on the outside of the parking lanes. Each drive was 14 km (8.7 miles) and took approximately 18 minutes to complete.

The secondary task provides a controlled manipulation of the demands that emerging in-vehicle technology places on drivers. The task required participants to listen to and respond to auditory

messages that were represented by a synthetic, English-speaking male adult (Reyes & Lee, 2004). Each message presented information on the cost (one dollar sign or two dollar signs), quality (one star or two stars), and wait time (short or long) for three different restaurants. After listening to each message, participants were asked to respond to six questions that required transforming the presented information to categories of restaurants.

Certain modifications were necessary when adapting the cue-target paradigm for the driving simulator environment: 1) instead of having a fixation point (Posner, 1980), participants drove through a natural scene and monitored a lead vehicle that braked periodically; 2) instead of monitoring a limited number of targets over a few seconds, participants were asked to scan a complex environment for many potential targets over a few minutes; and 3) instead of having one cue and one target presented serially for each trial, participants had to detect 20 targets after the onset of each cue. Given that the duration of each drive was approximately 18 minutes, participants' visual attention was less carefully controlled such that participants could have multiple fixations toward the potential target locations before detection. These modifications were done to expose participants to a complex dynamic situation that is more representative of everyday activities than is the traditional cue-target paradigm.

## **Procedure**

Participants drove a practice drive to get accustomed to the vehicle dynamics, driving environment, and the detection task. Participants also practiced the secondary task while sitting in the simulator. They were encouraged to engage in the secondary task with a 20-cent incentive for each correct answer. The main experimental drives began after participants indicated that they fully understood the instructions. The criteria of this assessment included whether or not the driver depressed the brake when the lead vehicle was braking, pressed a button upon seeing a pedestrian, and responded to the auditory questions in terms of restaurant names. Participants were told to scan the driving scene and drive as they normally would.

## **Experimental Design and Dependent Variables**

The study used a within-subjects design with the following factors: secondary task (task, no-task), scene clutter (high, low), and cue validity (valid, neutral, and invalid). The order of conditions on the three factors was counterbalanced to minimize learning effects. Secondary task and scene clutter varied between drives. Participants performed the secondary task in two experimental drives and confronted scene clutter in two experimental drives (one drive in the task condition and one drive in the no-task condition). Results discussed in this paper were from the two drives that had the secondary task. An equal number of the braking and pedestrian events occurred during the two phases of the secondary task.

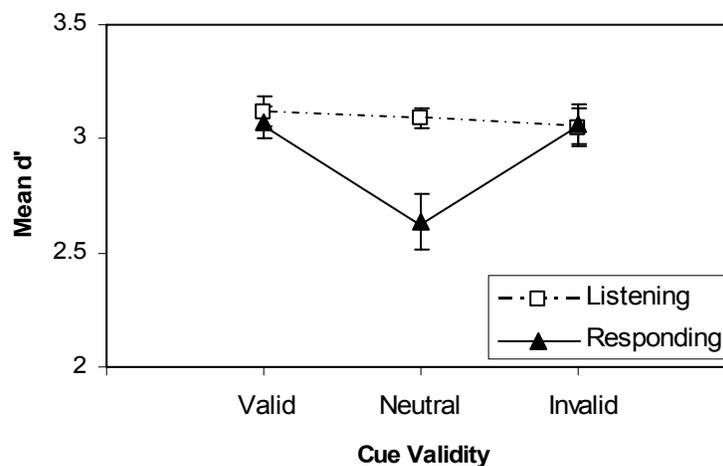
Cue validity varied within drives. There were four sections in each drive, and each section was defined by the configurations of signs (either one pedestrian crossing sign and one merge lane sign or two pedestrian crossing signs) that cued the 20 pedestrian targets. Upon seeing the cues, participants were required to report the location of the pedestrian crossing sign(s). This ensured that participants knew the side(s) of the roads that were cued by the pedestrian sign(s). For sections that began with one pedestrian crossing sign either on the right or left side of the road,

the sign conveyed predictive information about the location of the upcoming targets. The pedestrian crossing sign was analogous to the arrows in the cue-target paradigm (Jonides, 1981; Posner, 1980). The pedestrian crossing sign validly cued the location of 16 targets (80%) and invalidly cued the location of 4 targets (20%). Participants were informed that there would be more pedestrians on the same side as the sign, but they were not informed of the actual percentage. For sections that began with two pedestrian crossing signs, one on either side, the signs conveyed non-predictive information about the location of the upcoming events. The pedestrian crossing signs in this condition were analogous to a neutral double-headed arrow (Berger, Henik, & Rafal, 2005; Laubrock, Engbert, & Kliegl, 2005) or a diamond-shaped cue (Jonides, 1980) in the traditional cue-target paradigm. There were 10 events that occurred on the right side of the road and 10 events that occurred on the left side of the road; all were neutrally cued by the pedestrian crossing signs. The order of the pedestrian crossing signs was counterbalanced for each drive and across participants and the experimental conditions according to a Graeco-Latin square design.

Sensitivity ( $d'$ ) in the detection task, number of eye fixations during three-second response windows in pedestrian areas, and conditional probability of hit given a fixation in pedestrian areas during the listening and responding phases were analyzed.

## RESULTS

Phase of secondary task affected  $d'$  significantly,  $F(1,165) = 6.04, p = .015, d = 0.33$ . Participants were more sensitive to targets during the listening phase ( $M=3.09$ ) compared to the responding phase ( $M=2.92$ ), which were both lower than during the no-task condition ( $M=3.36$ ). Cue validity affected  $d'$  significantly,  $F(2,165) = 4.77, p = .010$ . The mean  $d'$  was comparable between valid events (3.10) and invalid events (3.06), and both were higher than the mean for neutral events (2.86). The interaction between phase and cue validity was significant,  $F(2,165) = 5.03, p = .008$ , such that responding to questions only affected participants' sensitivity in detecting neutrally-cued pedestrians (Figure 1). Scene clutter did not significantly decrease participants' sensitivity during either the listening or the responding phase,  $F(1,165) = 2.70, p = .102$ .



**Figure 1. Mean  $d'$  ( $\pm$ SE) as a function of cue validity and phase of secondary task**

Phase of auditory task affected number of eye fixations in the pedestrian areas,  $F(1,165) = 4.65$ ,  $p = .032$ ,  $d = 0.21$ , with a mean number of fixations of 5.55 during listening and 4.79 during responding. Cue validity affected number of fixations,  $F(2,165) = 10.81$ ,  $p < .0001$ , with a mean comparable between valid (5.89) and neutral events (5.59); both were higher than the mean for invalid events (4.02). Scene clutter did not affect number of fixations,  $F(1,165) < 1$ .

Phase of auditory task did not affect the probability of hit given a fixation in the pedestrian areas,  $F(1,165) = 3.36$ ,  $p = .069$ ,  $d = 0.24$ . The mean probability was .88 during the listening phase and .80 during the responding phase. Cue validity affected the probability of hit given fixation,  $F(2,165) = 11.41$ ,  $p < .0001$ . The mean probability was comparable between valid (.93) and neutral (.90) events, and both were higher than the mean for invalid events (.69). Scene clutter did not affect the probability of hit given fixation,  $F(1,165) < 1$ . The interaction between phase and clutter was significant,  $F(1,165) = 4.79$ ,  $p = .030$ , indicating that scene clutter decreased the probability of hit given fixation during responding, but increased the probability during listening.

## DISCUSSION

Drivers' guidance of attention was manipulated by having them follow cues that provided valid, neutral, and invalid information concerning target locations. During the responding phase, drivers fixated the target locations less frequently and were less sensitive to the appearance of targets compared to the listening phase. This finding is consistent with results of Strayer and Johnston (2001) and Levy et al. (2006), and suggests that drivers experience dual-task slowing when performing two concurrent tasks, each requiring response generation. The responding phase did not significantly affect the probability of a hit upon fixation on targets, indicating that information consolidation was not impaired by the voice interactions. Therefore, the dual-task slowing observed in the current study is mostly related to disrupted attention allocation, instead of information consolidation.

Drivers were less able to detect targets when they received non-predictive neutral cues. Instead of finding decreasing sensitivity with decreasing cue validity, as would be predicted by a traditional approach of cue-target paradigm, our results showed that valid and invalid cues provide a larger amount of information on target locations. The neutral cues convey little endogenous guidance of attention, and performance on another task interferes with the execution of attention allocation.

Although Kubose et al. (2006) found similar driving performance (e.g., control of velocity and maintenance of headway time) during listening and responding to auditory messages, our results suggest that voice interactions have the potential to impair drivers' attention allocation, especially when they can not rely on their endogenous control of attention.

## ACKNOWLEDGMENTS

The work presented here is part of the SAfety VEhicle(s) using adaptive Interface Technology (SAVE-IT) program that was sponsored by the U.S. DOT - National Highway Traffic Safety

Administration (NHTSA) (Project Manager: Michael Perel) and administered by the John A. Volpe National Transportation Systems Center (Project Manager: Mary D. Stearns).

## REFERENCES

- Alm, H., & Nilsson, L. (1994). Changes in driver behavior as a function of handsfree mobile phones - a simulator study. *Accident Analysis and Prevention*, 26(4), 441-451.
- Berger, A., Henik, A., & Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology-General*, 134(2), 207-221.
- Blincoe, L. J., Seay, A. G., Zaloshnja, E., Miller, T. R., Romano, E. O., Luchter, S., et al. (2002). *The economic impact of motor vehicle crashes 2000* (No. Report no. DOT HS-809-446). Washington, DC: US Department of Transportation.
- Glassbrenner, D. (2005). *Driver cell phone use in 2005 - Overall results*. Retrieved January 10, 2006, from <http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/RNotes/2005/809967.pdf>
- Ishihara, S. (1966). *Ishihara's tests for colour-blindness*. Tokyo, Japan: Isshinkai Foundation.
- Jonides, J. (1980). Towards a model of the mind's eye's movement. *Canadian Journal of Psychology-Revue Canadienne De Psychologie*, 34(2), 103-112.
- Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention & Performance IX*. Hillsdale, NJ: Erlbaum, 187-203.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-Car Naturalistic Driving Study data [Electronic Version]. Retrieved January 7, 2007 from <http://www-nrd.nhtsa.dot.gov/departments/nrd-13/810594/images/810594.pdf>.
- 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(1), 43-63.
- Laubrock, J., Engbert, R., & Kliegl, R. (2005). Microsaccade dynamics during covert attention. *Vision Research*, 45(6), 721-730.
- Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors*, 43(4), 631-640.
- Levy, J., Pashler, H., & Boer, E. (2006). Central interference in driving - Is there any stopping the psychological refractory period? *Psychological Science*, 17(3), 228-235.
- National Center for Injury Prevention and Control (2004). *WISQARS Leading Causes of Death Reports, 1999 – 2004*. Retrived April 4, 2007 from <http://webappa.cdc.gov/sasweb/ncipc/leadcaus10.html>
- Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention*. Hove, England: Psychology Press, 155-189.

- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.
- 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(2), 119-137.
- Reyes, M. L., & Lee, J. D. (2004). The influence of IVIS distractions on tactical and control levels of driving performance. In *Proceedings of Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society, 2369-2373.
- Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology-Applied*, 9(1), 23-32.
- 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(6), 462-466.
- US Department of Transportation (2004). *Traffic Safety Facts 2004*. Retrived April 4, 2007, from <http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/TSF2004.pdf>