

VERBAL-SPATIAL CUE CONFLICT: IMPLICATIONS FOR THE DESIGN OF COLLISION-AVOIDANCE WARNING SYSTEMS

Jane H. Barrow and Carryl L. Baldwin
George Mason University
Fairfax, Virginia, USA
Email: jbarrow1@gmu.edu

Summary: A spatial auditory Stroop paradigm was used to examine the effects of verbal-spatial cue conflict on response accuracy, reaction time, and driving performance. Participants responded to either the semantic meaning or the spatial location of a directional word, which were either congruent (i.e. the word “right” being presented from the right) or incongruent (i.e. the word “right” being presented from the left), while following a lead car in a simulated driving scenario. Accuracy was worse when participants were responding to the spatial location of a word in an incongruent trial, indicating that participants experienced significant interference when trying to ignore the semantic meaning of the word when it conflicted with the presentation location. Implications for the design of collision-avoidance warning systems are discussed.

INTRODUCTION

Automated systems are increasingly being implemented in modern automobiles in an effort to increase safety. Collision avoidance warning systems are an important component of this effort (Ho & Spence, 2005; Wang, Pick, Proctor, & Ye, 2007). The most effective means by which to alert drivers remains a point of debate, but research has shown that spatial auditory cues speed reaction time to stimuli and that multi-modal directional cues can speed reaction time even further (Bolia, et al., 1999; Gondan, et al., 2005; Tannen, et al., 2004). As a result, much research in this field has focused on the spatial nature and modality of warnings.

Because driving is a visually demanding task, the auditory modality is ideal for delivering effective collision warnings (Wickens, 1984). The relative benefit of alerting drivers to the relevant location of a hazard with spatial versus semantic audio cues remains a question of interest. Ho and Spence (2005) found that participants responded faster to verbal directional cues than they did to non-verbal directional cues, indicating that the semantic information provided by the verbal directional cue was processed more quickly than spatial information provided by the non-verbal directional cue. However, when the two cues were combined to create a congruent verbal-spatial directional cue, participants responded faster still. This finding indicates that having redundant information can actually speed reaction time (i.e. both verbal and non-verbal/spatial directional cues provide the same information), which is supported by research on multi-modal redundant targets (Bolia, et al, 1999; Gondan, et al., 2005; Tannen, et al., 2004). However, the study did not investigate the effect of an incongruent, or conflicting verbal-spatial directional cue.

Clearly, designers of collision avoidance warning systems would not intentionally use conflicting pieces of information to relay a warning to the driver; however, modern vehicles are not a silent

environment and warning systems are not 100% reliable. Many drivers utilize GPS navigation systems while driving, which provide verbal directions about when and where to make turns in their route. When directional information is being provided by two different sources, especially when these sources may be using different forms of auditory directional cues, there is an increased possibility that directional information from one source could conflict with directional information from another source. Wang and colleagues (2007) touched on this very issue in their investigation of driving responses to a Side Collision-Avoidance System (SCAS) when navigation signals were present. They found no differences in reaction time to the SCAS warning when the navigation signal corresponded with the SCAS warning and when it conflicted, but the navigation information was provided visually while the SCAS warning was provided aurally. Research using cross-modal Stroop paradigms has shown that when auditory and visual cues conflict, there is a significant lag in reaction time to the target when presented with an invalid auditory cue but not when presented with an invalid visual cue (Mayer & Kosson, 2004). This suggests that visual information is easier to ignore than auditory information, which could explain why there was no difference in reaction times for conflicting and non-conflicting cues from the navigation and SCAS systems in Wang and colleagues' (2007) study. Participants could have been prioritizing the auditory SCAS warning.

To further investigate this issue, the current study utilized a spatial auditory Stroop task originally used by Pieters (1981). The paradigm consists of verbal directional information presented from either a congruent spatial location (i.e. the word "right" presented from the right) or an incongruent spatial location (i.e. the word "right" presented from the left). Participants would either be responding to the spatial location of the stimulus, or the semantic meaning of the stimulus. Participants performed this task while following a lead car in a simulated freeway environment on a desktop driving simulator.

METHOD

Participants

Voluntary participation was obtained from 18 undergraduates (16 female) with a mean age of 19.69 years ($SD = 2.02$) enrolled in a university on the east coast. All participants reported normal or corrected-to-normal vision and passed an audiometric assessment of their hearing, indicating that their puretone threshold was less than 24 dB at 250-8000 Hz. All participants were fluent in English.

Materials and Apparatus

Auditory stimuli consisted of the words "right", "left", and "house" spoken in a naturalistic female voice, digitized and then presented in either the right channel, left channel, or both channels. All auditory stimuli were presented at a level approximating 60 dB from free field computer speakers. The speakers were placed 42 inches apart, with the participant seated directly between them.

The simulated driving task required participants to follow a lead car while maintaining a consistent headway, speed, and lane position on a four-lane freeway with no ambient traffic.

When the participant began driving, the lead car began to move forward, then sped up to maintain a constant speed of 65 mph. Images of common brand logos were presented on billboards on both sides of the road during the driving simulation. Two series of billboard images were constructed so that no images were repeated from one condition to the next.

Experimental Tasks and Design

Auditory tasks. Trials consisted of the words “right” or “left” coming from the right or left speaker. Stimuli were the same in the two auditory tasks, with the exception of control trials, but the instructions changed the nature of how the task was performed. Each task consisted of congruent, incongruent, and control trials as detailed below. Reaction time and accuracy were recorded for both tasks.

In the **semantic task**, participants were instructed to respond to the semantic meaning of the word by depressing a key representing “right” if they heard the word “right” and vice versa for the word “left”, regardless of the spatial location of the word. Congruent trials occurred when the semantic meaning of the word matched the presentation location (i.e. the word “right” came from the right), and incongruent trials occurred when the semantic meaning of the word did not match the presentation location (i.e. the word “right” came from the left). A control trial occurred when the word “right” or “left” came from both speakers, eliminating the directionality of presentation location.

In the **location task**, participants were instructed to indicate the spatial location of the word presented by depressing a key representing “right” if they heard a word presented from the right and vice versa for a word presented from the left, regardless of the semantic meaning of the word. A control trial in this task consisted of the word “house” coming from either the right or the left speaker, eliminating the semantic meaning of the spoken word in terms of directionality.

Driving task. Participants were instructed to follow the car in front of them at what they deemed to be a safe following distance, while maintaining a speed of 65 mph and their lane position. In the event that the participant lost the lead car (fell too far behind to safely catch up), they were instructed to maintain their speed and lane position, and not worry about trying to catch up to the lead car. Average speed and lane deviation were measured.

Billboard task. Participants were instructed to remember as many of the logos on the billboards as possible while performing the other two tasks. The experimenter clearly indicated that this was the lowest priority task – participants were asked to focus on maintaining their driving performance and their speed and accuracy on the auditory task. Participants received two scores: one for the number of correct, freely recalled logos, and one for the number of logos recognized in a subsequent recognition test that included both old and new logos.

Design. A 2x3 mixed-factorial design was used to examine the effects of response type (semantic vs. location) and congruency (congruent, control, or incongruent). Dependent measures were reaction time and accuracy for the auditory tasks, deviation from average speed and lane position for the driving task, and the number of correctly recalled and recognized logos for the billboard task.

It was hypothesized that performance on the dependent measures would be better in congruent trials than in incongruent trials, and that performance would also be better in the location auditory task than the semantic auditory task, based on the nature of the auditory system. It was also hypothesized that due to the predicted preference for responding to location information over semantic content of an auditory cue, incongruent trials where a participant was performing the semantic auditory task (and therefore ignoring location information) would result in poorer performance on the dependent measures.

Procedure

Upon entering the laboratory, participants were given an audiometric assessment and then completed a demographic questionnaire, way-finding surveys and the Edinburgh Handedness Inventory (Oldfield, 1971). For the first block of the experiment, the experimenter verbally gave instructions to the participant on how to perform the auditory task, allowed the participant to practice the task, and then gave instructions to the participant on how to perform the driving task, followed again by practice. The participant then practiced both tasks together. The experimenter gave verbal instructions on the billboard task, reiterated the instructions for the auditory and driving tasks, then started the experimental trials. At the end of the experimental trials, the participant completed the NASA-TLX (Hart & Staveland, 1988) with instructions to rate workload only on the auditory task. Next, the participant freely recalled the images that he or she remembered from the billboards, and then went through a slideshow of images to indicate which images they had seen in the driving scene and which were novel. The participant was offered a break, and then followed the same procedure for the second block of the experiment, minus the practice session for the driving task, since it did not change. The order of auditory tasks was counterbalanced across subjects, as were the driving scenes. Additionally, a baseline was taken of the participant's response time to each word in the auditory task (without the presence of spatial information). In half the participants, the baseline was taken prior to starting the first block of the experiment, and in the other half, the baseline was taken after the second block.

RESULTS

Two participants (both female) were excluded from the analyses due to computer failure during the experimental session, which resulted in incomplete data being recorded. Examination of the baseline data revealed that participants responded significantly faster to the word "house" than they did to either "right" or "left", $F(2,30) = 27.32$, $p < .05$, but that there was no difference in response time to the words "right" and "left". This observation indicates that the digitized word "house" may have been more acoustically salient, resulting in people consistently responding to it faster. We excluded all control trials from the analysis due to this confound, and only examined the differences between congruent and incongruent trials.

Auditory Tasks

Descriptive statistics for reaction time and accuracy to the auditory task trials can be found in Table 1. A two-way repeated measures MANOVA revealed that accuracy was better in the congruent trials than the incongruent trials, $F(1,15) = 18.23$, $p < .05$. Accuracy was also significantly better in the semantic condition than in the location condition, regardless of the

congruency of the trial, $F(1,15) = 13.13, p < .05$. This was interesting, since we originally predicted that people would be faster and more accurate in the location condition. Further analysis revealed an interaction between response type and congruency that approached significance, $F(1,15) = 2.98, p = .11$. The trend in the data revealed that accuracy was worse in the incongruent location trials than in the incongruent semantic trials, falling below 90% accuracy (see Figure 1). Reaction time in the congruent trials was significantly faster than in the incongruent trials, regardless of response type, $F(1,15) = 18.23, p < .05$. However, none of the remaining reaction time data approached significance, though they did show the same trends that were seen in the accuracy data.

Table 1. Descriptive Statistics for Auditory Tasks

	Condition	Trial Type	Mean	SD
Reaction Time (ms)	Location	Congruent	921.34	187.56
		Incongruent	973.39	198.65
	Semantic	Congruent	915.87	139.32
		Incongruent	948.70	147.79
Accuracy (Percentage correct response)	Location	Congruent	.96	.05
		Incongruent	.87	.10
	Semantic	Congruent	.98	.02
		Incongruent	.95	.05

Driving and Billboard Tasks

Driving data (average speed and lane deviation) and billboard logo recall and recognition were analyzed using two one-way repeated measures MANOVAs. Comparisons were only made between performance on the semantic auditory task and the location auditory task. No significant differences were observed for any of these measures.

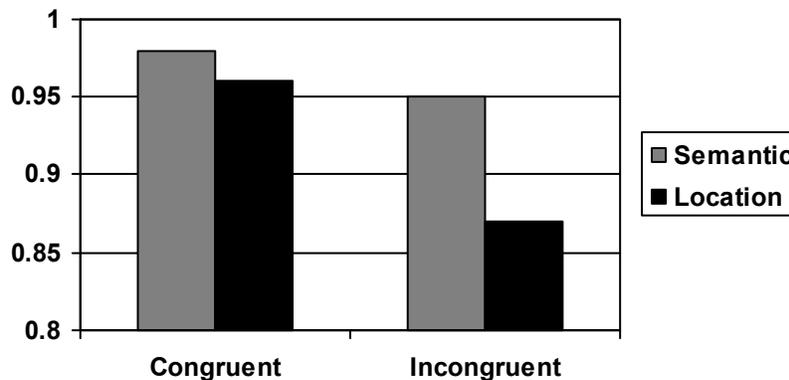


Figure 1. Accuracy for each response condition plotted as a function of congruency

DISCUSSION

The results of this study support those of Ho and Spence (2005), indicating that congruent verbal-spatial directional information leads to a faster response than non-spatial information. Additionally, verbal directional information results in a faster response than non-verbal

directional information. Accuracy data in the current study show that participants were more accurate when responding to verbal (semantic) information relative to when they were responding to non-verbal (location) information, and the reaction time data show a similar trend. Wang and colleagues (2007) found no difference in reaction time to an auditory collision avoidance warning whether it conflicted with visually presented navigation directions or not. In the present study, we did not manipulate congruence of visual stimuli but focused entirely on the congruence of auditory information. Our results indicate a marginally significant difference in accuracy for incongruent trials depending on whether semantic or location information was being attended. Specifically, participants were less accurate when responding to the location of a word, and therefore trying to ignore conflicting semantic information. This further supports Ho and Spence's (2005) results that show the importance of verbal directional cues over non-verbal spatial directional cues in terms of improving performance – the decrease in accuracy demonstrates that attempting to ignore the semantic content of a conflicting stimulus was more difficult than attempting to ignore the spatial location.

This finding has important implications for the design of collision avoidance warning systems. While the results of this and other studies (Ho & Spence, 2005) support the benefit of having a congruent verbal-spatial directional cue for alerting drivers to an impending collision, there is the potential for a huge cost. If the spatial directional information from the collision avoidance system were to conflict with other verbal directional information present in the vehicle, these results predict that making the correct response to the collision avoidance warning would be more difficult, possibly resulting in the wrong response. Is the added benefit of the congruent verbal-spatial directional cue great enough to outweigh the cost of an increased error rate? Is the added cost of incorporating spatial audio capabilities into these types of systems worth it when manufacturers could just as easily relay information with non-spatial verbal cues? This type of cost-benefit analysis can't be answered with research, but would hopefully be examined by the designers of these systems.

ACKNOWLEDGMENTS

The authors would like to thank Dan Roberts and Tony Layton for their assistance with data collection, and their feedback regarding the experimental tasks and experimental set-up. The authors would also like to thank David Kidd for his invaluable help in programming the driving scenarios.

REFERENCES

- Bolia, R.S., D'Angelo, W.R., & McKinley, R.L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors*, 41, 664-669.
- Gondan, M., Neiderhaus, B., Rosler, F., and Roder, B. (2005). Multisensory processing in the redundant-target effect: A behavioral and event-related potential study. *Perception & Psychophysics*, 67(4), 713-726.
- Hart, S.G. & Staveland, L.E. (1988). "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." In P.A. Hancock & N. Meshkati (eds), *Human Mental Workload* (239-250). Amsterdam: North Holland Press.

- Ho, C & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied*, 11, 157-174.
- Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113.
- Mayer, A.R. & Kosson, D.S. (2004). The effects of auditory and linguistic distractors on target localization. *Neuropsychology*, 18, 248-257.
- Pieters, J.M. (1981). Ear asymmetry in an auditory spatial Stroop task as a function of handedness. *Cortex*, 17, 369-380.
- Tannen, R.S., Nelson, W.T., Bolia, R.S., Warm, J.S., & Dember, W.N. (2004). Evaluating adaptive multisensory displays for target localization in a flight task. *The International Journal of Aviation Psychology*, 14, 297-312.
- Wang, D., Pick, D.F., Proctor, R.W., & Ye, Y. (2007). Effect of a side collision-avoidance signal on simulated driving with a navigation system. *Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Portland, Oregon, 206-211.
- Wickens, C.D. (1984). "Processing resources in attention." In R. Parasuraman and R. Davies (eds), *Varieties of attention* (63-101). New York: Academic Press.