SPATIAL AUDIO VS. VERBAL DIRECTIONAL CUES: AN EXAMINATION OF SALIENCE AND DISRUPTIVENESS WITHIN A SIMULATED DRIVING CONTEXT

Jane H. Barrow and Carryl L. Baldwin

George Mason University
Psychology Dept., Arch Lab
4400 University Dr., MS 3F5, Fairfax, VA 22030, USA
jbarrow1@gmu.edu and cbaldwi4@gmu.edu

ABSTRACT

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 was 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. Greater performance decrements were observed when participants were attempting to ignore a semantically incongruent verbal cue relative to incongruence from the spatial location of the cue. Implications for the design of spatial auditory displays are discussed.

1. INTRODUCTION

Spatial auditory cues and verbal directional cues are both viable options for alerting the human operator to danger in a variety of environments [1], [2], [3], [4]. The question of interest here is which type of information is more salient, and therefore more disruptive when irrelevant or erroneous. Understanding the roles that verbal and spatial components of an auditory message play is an important step in learning to design cues and alerts for a variety of display applications, not just for alerts to potential dangers in the environment.

For the purposes of this paper, the verbal and spatial location components of auditory alerts are examined in terms of their spatial orienting effectiveness within the driving domain. Automated systems are increasingly being implemented in modern automobiles in an effort to increase safety, and because driving is a visually demanding task, the auditory modality is ideal for presenting supplementary information to aid the driver [5]. There is not yet a consensus in the literature as to the relative benefits of alerting drivers to the relevant location of critical events with spatial versus semantic audio cues. Ho and Spence [2] 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). This observation is supported by research on multi-modal redundant targets [3], [6], [7]. However, Ho and Spence did not investigate the effect of an incongruent, or conflicting verbal-spatial directional cue.

Clearly, designers would not intentionally use conflicting pieces of information to relay spatial information to the driver. However, technologies are not infallible and modern vehicles are not a silent environment. 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, Pick, Proctor, and Ye [4] 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 [8]. 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 et al.'s 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 [9]. 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. This allowed us to examine the relative interference of the spatial component of the stimulus and the semantic meaning of the stimulus.

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.

2. METHOD

2.1. 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 hearing level was less than 24 dB across 250-8000 Hz. All participants were fluent in English.

2.2. 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.

2.3. Experimental Design and Tasks

2.3.1. 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.

2.3.2. 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.

2.3.3. 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.

2.3.4. 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.

2.4. 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 [10]. 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 [11] 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.

	Condition	Trial Type	Mean	SD
RT (ms)	Location	Congruent	921.34	187.56
		Incongruent	973.39	198.65
	Semantic	Congruent	915.87	139.32
		Incongruent	948.70	147.79
%Correct (percentage correct)	Location	Congruent	.96	.05
		Incongruent	.87	.10
	Semantic	Congruent	.98	.02
		Incongruent	.95	.02

Table 1: Descriptive statistics for auditory tasks.

3. 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.

3.1. Auditory Tasks

Descriptive statistics for reaction time and accuracy to the auditory task trials can be found in Table 1. As predicted, 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,

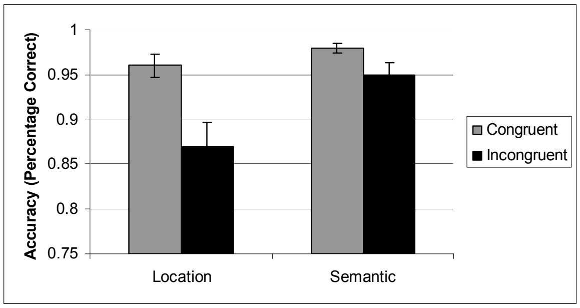

Figure 1: Accuracy for each response condition plotted as a function of congruency. Error bars represent the SEM.

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 indicated that incongruent semantic information tended to disrupt performance to a greater degree than incongruent location information (see Figure 1 and 2). Both reaction time and accuracy suffered to a greater degree from the incongruent semantic information. This suggests that the overall superior performance in the semantic condition may have resulted primarily from the absence of detrimental effects of incongruent spatial location information.

3.2 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.

4. DISCUSSION

The results of this study support those of Ho and Spence [2], 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 et al. [4] 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 demonstrated a trend for greater interference from incongruent semantic information when responding to the location of a word, relative to incongruent spatial information when responding to the physical location of a sound. This further supports Ho and Spence's [4] results that show the importance of verbal directional cues relative to non-verbal spatial directional cues in terms of improving performance. However, these results raise an important caveat. Depending on the reliability of the system, using semantic methods of presenting spatially predictive information may pose significant problems. Incongruent semantic information may cause greater disruption. It may be more difficult to ignore the semantic content of a conflicting stimulus rather than its spatial location, thus potentially negating the benefit of semantic spatially predictive cues.

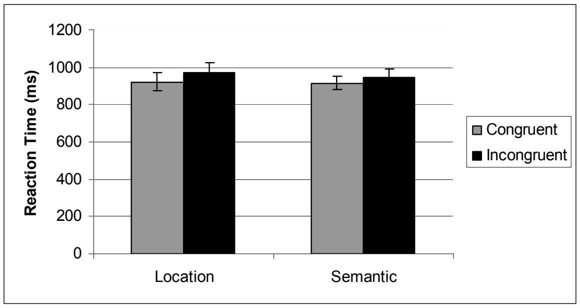

Figure 2: Reaction time for each response condition plotted as a function of congruency. Error bars represent the SEM.

These findings support previous research demonstrating the salience of semantic information, but also illustrate the potential for that information to disrupt response to other tasks. In imperfect systems, it might be wiser to use spatial audio information, which is valuable in directing attention, and less disruptive, particularly in situations where the other tasks being performed require directional judgments and may be equally important to the nature of the spatial audio alert.

5. 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.

6. REFERENCES

[1] D.R. Begault, "Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation," in *Human Factors*, vol. 35, no. 4, pp. 707-717, 1993.

[2] C. Ho & C. Spence, "Assessing the effectiveness of various auditory cues in capturing a driver's visual attention," in *Journal of Experimental Psychology: Applied*, vol. 11, pp. 157-174, 2005.

[3] R.S. Tannen, W.T. Nelson, R.S. Bolia, J.S. Warm, & W.N. Dember, "Evaluating adaptive multisensory displays for target localization in a flight task," in *International Journal of Aviation Psychology*, vol. 14, pp. 297-312, 2004.

[4] D. Wang, D.F. Pick, R.W. Proctor, & Y. Ye, "Effect of a side collision-avoidance signal on simulated driving with a navigation system," in *Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Portland, Oregon, 2007, pp. 206-211.

[5] T.A. Dingus, M.C. Hulse, & W. Barfield, "Human-system interface issues in the design and use of Advanced Traveler Information Systems," in W. Barfield & T.A. Dingus (Eds), *Human Factors in Intelligent Transportation Systems*, Mahwah, NJ: Lawrence Erlbaum, 1998.

[6] R.S. Bolia, W.R. D'Angelo, & R.L. McKinley, "Aurally aided visual search in three-dimensional space," in *Human Factors*, vol. 41, pp. 664-669, 1999.

[7] M. Gondan, B. Neiderhaus, F. Rosler, & B. Roder, "Multisensory processing in the redundant-target effect: A behavioral and event-related potential study," in *Perception & Psychophysics*, vol. 67, no. 4, pp. 713-726, 2005.

[8] A.R. Mayer & D.S. Kosson, "The effects of auditory and linguistic distractors on target localization," in *Neuropsychology*, vol. 18, pp. 248-257, 2004.

[9] J.M. Pieters, "Ear asymmetry in an auditory spatial Stroop task as a function of handedness," in *Cortex*, vol. 17, pp. 369-380, 1981.

[10] R.C. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," in *Neuropsychologia*, vol. 9, no. 1, pp. 97-113, 1971.

[11] S.G. Hart & L.E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in P.A. Hancock & N. Meshkati (Eds), *Human Mental Workload,* Amsterdam: North Holland Press, 1988.