MULTISENSORY CUE CONGRUENCY IN THE LANE CHANGE TEST

Yuanjing Sun[1], *Jaclyn Barnes*[2], *Myounghoon Jeon*[1,2]

Mind Music Machine Lab
Michigan Technological University,
[1]Department of Cognitive and Learning Sciences, [2]Department of Computer Science,
1400 Townsend Dr. Houghton, MI 49931 USA
{ysun4, jaclynb, mjeon}@mtu.edu

ABSTRACT

Drivers interact with a number of systems while driving. Taking advantage of multiple modalities can reduce the cognitive effort of information processing and facilitate multitasking. The present study aims to investigate how and when auditory cues improve driver responses to a visual target. We manipulated three dimensions (spatial, semantic, and temporal) of verbal and nonverbal cues to interact with visual spatial instructions. Multimodal displays were compared with unimodal (visual-only) displays to see whether they would facilitate or degrade a vehicle control task. Twenty-six drivers participated in the Auditory-Spatial Stroop experiment [1] using a lane change test (LCT). The preceding auditory cues improved response time over the visual-only condition. When dimensions conflicted, spatial (location) congruency had a stronger impact than semantic (meaning) congruency. The effects on accuracy was minimal, but there was a trend of speed-accuracy trade-offs. Results are discussed along with theoretical issues and future works.

1. INTRODUCTION

For decades, in-vehicle technologies have rapidly increased. Given that vision is fully occupied while driving, technologies using other modalities, such as speech recognition or vibrotactile notifications, have become pervasive in vehicles. Multiple Resource Theory (MRT) [2] indeed supports the use of multimodal interfaces in the vehicle context. However, despite potentially allowing drivers to process more information in parallel, multimodal interfaces still occupy attentional resources. Does more information always mean more facilitation? With bad design, multimodal displays might cause information overload or degrade performance, which could lead to safety hazards on the road. For example, suppose that the personal navigation device (PND) tells a driver to make a left turn, but at the same time, the collision warning system alerts the driver that there is a hazard coming from the left lane. How would the driver respond to this conflicting information? Even though multimodal displays might benefit a single task, they might not always benefit multiple tasks, especially when modalities conflict with one another at the same time. The present study aims to address this issue to identify underlying mechanisms and provide design guidelines for in-vehicle multimodal-visual and auditory-displays.

1.1. Multiple Resource Theory

Wickens' MRT [3] has been used to predict or analyze interference between concurrently perceived signals. Two tasks that demand *separate* levels (e.g., one visual and one auditory tasks) will interfere with each other *less* than two tasks that both demand one level of a given dimension (e.g., visual and visual tasks). It provides a basic theoretical endorsement to the blooming implementation of multimodal interfaces. However, MRT is also challenged by multisensory illusions, such as McGurk illusion or Ventriloquist Illusion [4], where information from different channels are synthesized into a new, distinct signal. The conflict between MRT and multisensory illusion prompts a more detailed examination of how multisensory perception influences information processing.

1.2. Multimodal Benefits

Multimodality provides synergy at the cost of significantly less cognitive effort than processing information from a single modal channel [5]. By providing processing advantages for grouping and organizing signals with the lowest workload, redundancy in multimodal display can increase the bandwidth of concurrent information processing. Here, the arrangement of multimodal signals becomes decisive to the occurrence and strength of multimodal benefits. The degree of multimodal benefits follows both (1) spatial rules and (2) temporal rules. However, several conflicting studies make it difficult to identify exactly from where the facilitation derives.

1.2.1. Spatial rules

In his review of crossmodal spatial attention [6], Spence proposed the performance benefit on ipsilateral (on the same side) cued trials over contralateral (on the opposite side) cued trials. A possible mechanism for this might be "spatial proximity" between stimulus and response. In other words, a spatially predictive auditory or visual cue would always lead to an exogenous attentional shift and narrow down spatial attention to the cue direction. A spatially corresponding mapping of left stimuli to left responses and right stimuli to right responses yielded better performance (i.e., faster reactions and fewer errors) than a spatially incongruent mapping [7].

https://doi.org/10.21785/icad2017.015

1.2.2. Temporal rules

There are divergent research outcomes on the temporal interval between auditory cue and visual target. For example, crossmodal synesthesia [8] predicts a synchrony benefit. It claims that the responses to multimodal cues will benefit when there is a maximum overlap between cue and target. In contrast, Posner's spatial cuing task proposes a "preparation function", suggesting that the response time would become fastest when a priming tone was 200 ms ahead of the visual target [9]. In this line, the present study selected 200 ms as preceding timing as the asynchrony condition to contrast with the synchrony condition.

1.3. Type and Demand of Visual Tasks

Multimodality does not always provide benefits over unimodality. Sinnett, Soto-Faraco and Spence [10] manipulated perceptual load (frequency of visual targets) and working memory load (numbers of responses) to compare the redundant gain of multimodality. The result indicated that both multisensory facilitation and inhibition can be demonstrated by changing the task type and visual demand.

In particular, Wickens and other researchers [11] suggested that a redundant auditory display may facilitate a visual scanning task but not an ongoing visual tracking task. In audiovisual redundancy studies, ongoing visual tracking tasks require continuous visual attention. In the context of visual tracking tasks, there are periodic interrupting tasks that are discrete in nature. A meta-analysis of 29 studies [12] comparing visual-auditory tasks with visual-visual tasks has shown that auditory presentation for a discrete task resulted in a significant 15% performance advantage over visual-only presentation. In particular, the auditory advantage increased when the two visual inputs were end-to-end. In other words, the auditory cues were more helpful when the interval between two visual inputs was shorter (i.e., visual perceptual load is high). It can also be inferred that the auditory-visual facilitation would occur in visually-demanding tasks (e.g., the demand of the visual scanning task is higher than the visual tracking task). The lane change test includes both visual scanning (identifying a visual target) and visual tracking tasks (maintaining lane position). We anticipate the use of auditory cues will be more helpful for the visual scanning task than the visual tracking task.

1.4. Auditory-Spatial Stroop Task

Inheriting from the original color-word naming Stroop paradigm, researchers utilized the Auditory-Spatial Stroop task to investigate location-meaning conflicts in multimodal processing. Auditory-Spatial Stroop task, originally introduced by Pieters [13], consists of directional verbal cues presented congruently or incongruently with a visual target. Mayer and Kosson showed that there was a significant lag in reaction time (RT) to the target location when incongruent auditory cues were presented. However, incongruent visual cues did not delay RT to auditory targets [14]. It suggested that a visual distractor is easier to ignore than an auditory distractor. The asymmetric anti-distraction feature between vision and audition indicates that the modality of message in multitask signaling could interfere with the priority level in response selection. Barrow and Baldwin [15], [16] used the Auditory-Spatial Stroop task to simulate the potential location-meaning conflict that might happen under several multimodal in-vehicle devices (e.g., side collision avoidance warning and PND). For example, the word, "left" or "right" is presented in a congruent or incongruent position with its meaning. They found that it is more difficult to ignore the spatial information of the verbal cues than the semantic information when there was a location-meaning conflict.

2. THE CURRENT STUDY AND HYPOTHESES

Understanding different mechanisms involved in multisensory perception is important for choosing appropriate modalities to convey messages for certain tasks. Designers need to have an overall consideration of the implementation environment and priority schedule of all the tasks. The present study intends to ascertain the decisive mechanism(s) in multisensory perception. Since the interference in spatial, semantic, and temporal dimensions is not always orthogonal, the interference of the three dimensions was respectively compared with the visual-only condition. In the view of this research purpose, we constructed three major sets of hypotheses:

- Hypotheses 1: Spatial Rules
  - H1a: Spatially congruent audio-visual (A-V) pairs will have shorter RT than the visual-only condition.
  - H1b: Spatially incongruent A-V pairs will have longer RT than the visual-only condition.
  - H1c: If two above are true, it could be inferred that spatially congruent A-V pairs will have shorter RT than spatially incongruent A-V pairs.
- Hypotheses 2: Temporal Rules
  - H2a: Asynchronous (i.e., preceding auditory cues) A-V pairs will have shorter RT than the visual-only condition.
  - H2b: Synchronous A-V pairs will not have longer RT than the visual-only condition.
- Hypotheses 3: Spatial-Semantic Conflict
  - H3a: Spatiality will have a stronger impact than semanticity. Spatially incongruent and semantically congruent conditions will have longer RT than the visual-only condition.
  - H3b: Spatially congruent and semantically incongruent conditions will have shorter RT than the visual-only condition.

3. METHOD

3.1. Participants

Twenty-six participants (23 male, 3 female; M_{age} = 20.6, SD_{age} = 2.3; $M_{\text{YearsOfDriving}}$ = 4.5, $SD_{\text{YearsOfDriving}}$ = 2.86) were recruited from the undergraduate participant pool of an American technical university. Participants were native English speakers at least 18 years old. To control for driving skill, participants were required to possess a valid driver's license and have at least 2 years of driving experience.

	Nonverbal Cue		*Verbal Cue*			
Spatial	*Congruent*	*Incongruent*	*Congruent*		*Incongruent*	
Semantic	*N/A*	*N/A*	*Congruent*	*Incongruent*	*Congruent*	*Incongruent*
Synchronous	Track 1	Track 2	Track 3	Track 4	Track 5	Track 11
Asynchronous	Track 6	Track 7	Track 8	Track 9	Track 10	Track 12

Table 1: Summary of audio-visual stimulus mappings in spatial, semantic, and temporal dimensions for various tracks used in the experiment. Tracks had 78% intended cues and 22% distractor cues to prevent participants from anticipating actions based on the audio cues.

3.2. Stimuli

3.2.1. Visual Stimuli

Each track began with a START sign, then 18 sets of lane change signs, and ended with a FINISH sign. The lane change signs appeared in an overhead position on a gate or bridge over the simulated roadway. They were composed of one down arrow and two Xs in three separate black boards (shown in Figure 1).

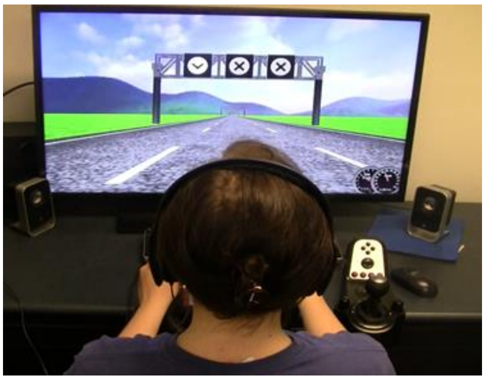

Figure 1: Visual target stimulus used in OpenDS lane change test scenarios.

3.2.2. Auditory Stimuli

Two non-verbal stimuli and four verbal stimuli were used as auditory cues in twelve tracks out of fourteen in total. The two non-verbal stimuli were a single and a double beep, indicating a single or double lane change. The non-verbal stimuli had no semantic congruency, but had spatial and temporal congruency.

The four verbal cues were “LEFT”, “RIGHT”, “LEF-LEFT”, and “RIGH-RIGHT”. The auditory stimuli were normalized to equal duration of 350ms at 60 dB level. The length and loudness of auditory cues were determined by reference to similar demands of the perceptual-motor experiments conducted in previous research [17]. All auditory stimuli were presented at a level of approximately 60 dB from the JVC-HA/RX300 stereo headset. The speech clips “LEFT” and “RIGHT” were recorded using the free online Text-to-Speech (TTS) service (www.naturalreaders.com) at medium speed with a female voice (Laura, US English). Sped up verbal clips “LEF-LEFT” and “RIGH-RIGHT” indicated the direction of double lane changes (i.e., from the left most lane of the three-lane simulated road to the right most lane or vice versa). These clips were created by importing the original TTS files “LEFT” to Audacity 2.1.0 and replicated the word “LEFT” to two audio tracks. For the first audio track, the first part “LEF” was kept and for the second audio track, the full word “LEFT” was kept. Finally, the two audio tracks were combined and compressed to 350 msec. The “Change Tempo” effect in Audacity was used to adjust the length of audio clip without changing the pitch.

In addition to temporal congruency, verbal cues had spatial congruency and semantic congruency. Thus, the mapping relationship of verbal cues with visual targets had both spatial congruency (physical location of the verbal cue to visual indication) and semantic congruency (meaning of the verbal cue to visual indication). For example, consider a semantically congruent and spatially incongruent condition. Given the visual cue for a single lane change to the left, the participant hears a verbal cue “LEFT” coming from the right speaker.

3.3. Driving Scenario and Apparatus

The Auditory-Spatial Stroop experiment was developed on the basis of the embedded ReactionTest scenario in OpenDS 2.5. We re-implemented the Lane Change Test Toolkit in OpenDS and made modifications according to ISO26022-2010. The original Lane Change Test (LCT) [18] is a simple laboratory dynamic dual task method, which quantitatively measures performance degradation in a primary driving task. In this way, researchers can manipulate the timing and multimodal combinations of lane change signs to capture different driving patterns under different conditions. The simulator consisted of SimuRide software with a 39” monitor and steering wheel. Speed was fixed and the pedals were not used. The primary task required a participant to drive in a straight three-lane road containing a series of lane changes defined by visual targets (Figure 1). In the original LCT, the simulated track length is 3,000 m, corresponding to three minutes of driving at a constant 60 kph [19]. However, to increase the perceptual workload in the primary driving task, the speed in the current experiment was increased to 110 kph (70 mph), a freeway speed limit in some US states. The 18 lane change signs were distributed at intervals of approximately 150 meters. In other words, each lane change maneuver needed to be completed within roughly 4 seconds. The lane change signs were made visible approximately 40 meters before the sign position. In this way, the lane keeping maneuver distinguished two successive lane change maneuvers and provided a buffer if participants made an erroneous lane change at the previous sign. The deviated

Order	Track Number													
1	0	8	2	11	6	4	7	1	10	13	5	12	3	9
2	0	9	3	12	5	10	1	7	4	13	6	11	2	8
3	0	7	4	6	12	2	8	9	3	13	5	11	10	1
4	0	1	10	11	5	3	9	8	2	13	12	6	4	7

Table 2: Summary of the partially counterbalanced exposure orders for the four participant groups. Visual-only baseline tracks 0 and 13, which remained constant for all groups, are highlighted in grey.

distance from the last sign did not influence the start position of the upcoming sign.

3.4. Experimental Design

Nonverbal tracks had two dimensions: spatial and temporal. Since the visual target appeared in every track, it was the reference for "congruent" or "incongruent". In the spatial dimension, a condition is congruent if the cue is played from the same side the driver should move toward, e.g., a single tone from the left side when the visual cue also indicates a single lane change to the left. In the temporal dimension, the asynchronous condition meant that the auditory cues appeared 200 ms ahead of the visual target. The synchronous condition indicated no temporal gap between audio-visual stimuli.

Verbal cues had three dimensions, spatial and temporal plus an added semantic dimension. Congruency in the semantic dimension refers to a match between the meaning of the word(s) in the cue and the desired maneuver.

The experiment was a 2 (spatial congruency) * 2 (semantic congruency) * 2 (temporal congruency) within-subjects design. All conditions are shown in Table 1. Apart from the twelve conditions, participants were given two chances of the baseline (visual-only) tracks, separately numbered as Track 0 and Track 13. The Track 13 were inserted between the 9th and 11th run to see the trend of the learning effect. Aside from the visual-only tracks, each track included 78% target cues and 22% distractor cues to prevent participants from anticipating maneuvers from the auditory cues. The order of 14 tracks was partly counterbalanced as shown in Table 2. Participants were randomly distributed into four groups. Orders 1 & 2 were reversed sequential orders. Order 3 split the tracks in the middle to the two extremes. Order 4 was the reversed sequence of Order 3. In this way, the order effects were minimized. To reduce participants' adaptation to repeated patterns, asynchrony, congruency, and modality were considered in each order.

3.5. Procedure

After signing a consent form, participants watched an instructional video for an overview of the experiment and guidance on how to maneuver the lane change test. The primary task in the LCT was to rapidly change lanes as directed by the visual targets and to maintain the center of the lane between maneuvers. At the same time, unpredicted auditory cues were sent out via the headset. The participants were required to count all auditory cues based on their locations (either left or right ear) and reported the subtotal number of each side to the experimenter at the end of each track.

Before the experiment started, an equivalent hearing test and training trial were given to the participant to make sure that all cues were recognizable to all participants. Also, the experimenter ensured that all participants comprehended the tasks in the whole process of the experiment. A RT histogram displayed briefly after the completion of each track. As long as the participant reached 50% accuracy, they were considered qualified to enter the formal portion of the experiment.

3.6. Metrics

Reaction time (RT) and percentage of correct lane changes (PCL) were two direct metrics for speed and accuracy [11]. The car position parameters (i.e., positional coordinates) were automatically recorded by the driving simulator at the sampling rate of 10 Hz [19]. The reaction to the stimulus was measured as the time span between stimulus and a steering wheel angle outside of the ordinary range for lane keeping. The reaction timer was activated simultaneously with the earlier cues' appearance and ceased when the car maintained the targeted lane for 800 ms. The 800 ms was then subtracted from the reaction timer's value, leaving the true reaction time as the output. The maximum RT window for correct completion of a lane change was either 4.1 seconds or 117 meters after the lane change sign (OpenDS Reaction Task default settings). Otherwise, it was recorded as an incomplete lane change. The reaction timer also excluded overshooting the target lane from recordings of correct lane change maneuvers.

The accuracy was the percentage of correct lane change (PCL) in each track. The correctness of lane change was defined by the driver's position before and after the lane-change maneuver [20]. For each road segment between two signs, the lane where the vehicle was most frequently positioned was identified. Consistent lane choices were then defined as those cases where the vehicle remained in the lane for more than 75% of the segment. This selected lane was then compared to the correct target lane. For each track, the PCL was then calculated as the fraction of the consistent lane

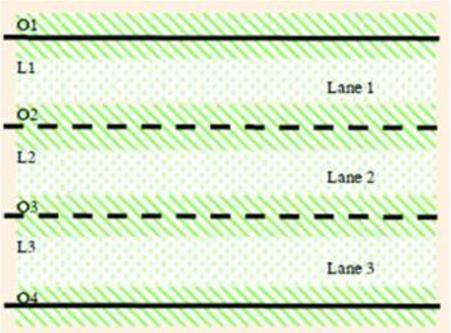

Figure 2: A diagram of effective area for reaction timer to distinguish correct lane change maneuver from erroneous or no lane change in LCT scenario [20].

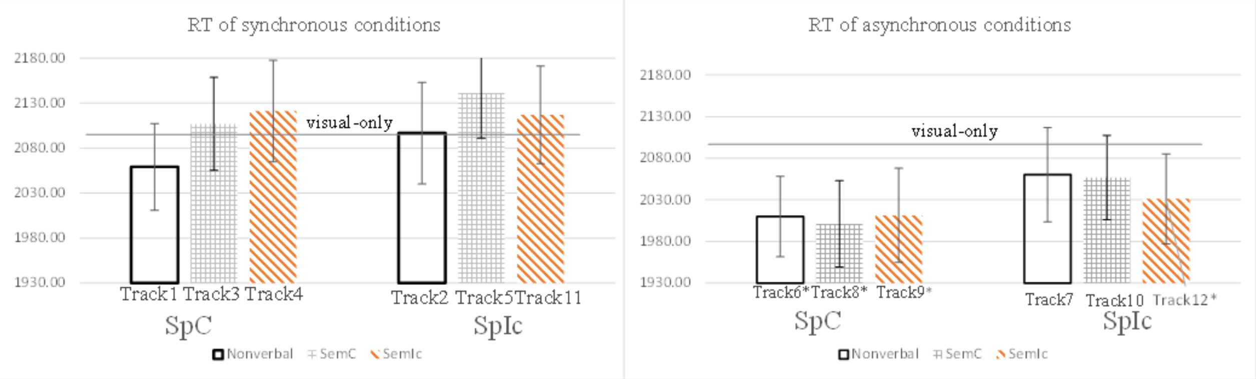

Figure 3: The left half plot is the average RT of synchronous conditions and the right half plot is the average RT of asynchronous conditions. Visual-only is marked as a baseline. Abbreviations used in the graph include synchronous (Syn), asynchronous (Asyn), spatial (Sp), semantic (Sem), congruent (C), and incongruent (Ic).

choices that were correct.

To determine this position, the 3-lane road was divided into different zones, corresponding to parts of the lanes as Figure 2 shows. The dotted zones L1 to L3 corresponded to a correct position in lane 1 (left lane), lane 2 (center lane) or lane 3 (right lane), while the stripe zones "O" corresponded to out of lane positions. The lateral position of the driver was defined by the zone which contained the 75% of his/her trajectory between two signs. If not, then the position was considered as being out of lane and the reaction timer outputted an NA instead of RT. The correctness of each lane change was defined as follows: (1) "Correct LC": the end position of the driver was in the intended lane; (2) "No LC": the driver was in the same Li zone at start and end positions; and (3) "Erroneous LC": the end position of the driver was in a different lane than both the starting lane and the target lane.

4. RESULTS

For planned comparisons, familywise Type I error rate is generally deemed unnecessary [21]. Thus, Bonferroni correction was not applied to the alpha level in the following paired samples t-tests. Twelve paired samples t-tests on RT and accuracy were respectively conducted to examine the mean difference between each condition track and the visual-only condition (mean of the two visual-only tracks).

Figure 3 shows average RT of correct lane-changes across all conditions with standard error bars. The visual-only condition is the baseline to mark facilitation versus deterioration. The unit of y axis is milliseconds. The asterisks in Tracks 6, 8, 9, & 12 show significant differences in paired samples t-tests when compared to the visual-only condition. For tracks with nonverbal cues, the asynchronous spatially congruent condition (Track 6) $t(26) = -2.383$, $p = 0.025$ showed significantly faster RT than the visual-only condition. For tracks with verbal cues, the asynchronous spatially congruent semantically congruent condition (Track 8) $t(25) = -2.478$, $p = 0.02$, the asynchronous spatially congruent semantically incongruent condition (Track 9) $t(25) = -2.817$, $p = 0.009$, and the asynchronous spatially incongruent semantically incongruent condition (Track 12) $t(25) = -2.665$, $p = 0.013$ showed significantly faster RT than the visual-only condition.

Figure 4 shows average accuracy in 12 conditions. For accuracy, there was no clear results or patterns, but synchronous conditions tended to show higher accuracy than asynchronous conditions.

Since the visual-only condition served as the baseline in comparison with all conditions, the subtraction of multimodal tracks over the visual-only tracks are denoted as ΔRT and Δ% in RT and accuracy respectively between multimodal tracks and visual-only tracks. This simplified version of the twenty-four paired samples t-test results is used in the discussion.

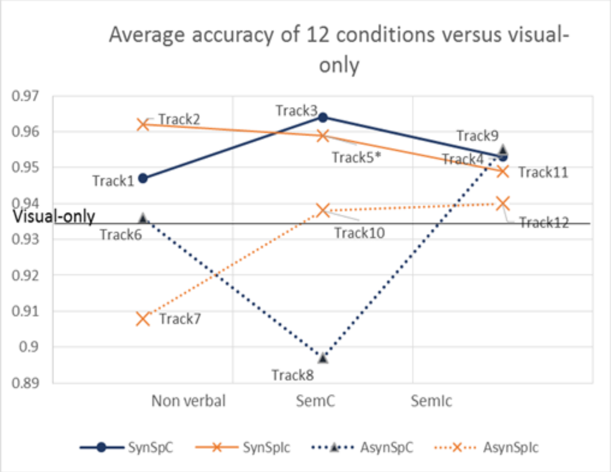

Figure 4: Average accuracy of 12 conditions versus visual-only. Abbreviations used in the graph include synchronous (Syn), asynchronous (Asyn), spatial (Sp), semantic (Sem), congruent (C), and incongruent (Ic).

5. DISCUSSION

The present experiment used the Auditory-Spatial Stroop paradigm [13] in a lane change test scenario to measure the variance of driving performance under the manipulation of spatial, semantic, and temporal congruency of auditory and visual cues.

5.1. Spatial Rules (H1)

The results showed that spatially congruent conditions, at least in the asynchronous conditions (Tracks 6, 8, & 9), had significantly faster RT than the visual-only condition. This partially supported H1a. It demonstrated that spatially congruent A-V association enhances visuospatial response speed. As with the spatial rules in multimodal facilitation, it

		Nonverbal Cue		*Verbal Cue*			
	Spatial	*Congruent*	*Incongruent*	*Congruent*		*Incongruent*	
	Semantic	*N/A*	*N/A*	*Congruent*	*Incongruent*	*Congruent*	*Incongruent*
Synchronous	*ΔRT*	-32.31	5.28	15.51	50.06	29.82	25.46
	Δ%	1.20%	2.70%	2.90%	1.80%	2.50%	1.40%
Asynchronous	*ΔRT*	-81.64*	-31.17	-90.16*	-80.11*	-34.69	-60.27*
	Δ%	0.10%	-2.70%	-3.70%	2.00%	0.30%	0.50%

Table 3: Subtraction of conditional RT and accuracy out of baseline RT and accuracy. Asterisks indicate statistical significance.

is easier to direct attentional focus in different sensory modalities to the same spatial location rather than different location [22]. However, the mixed results in the spatially incongruent conditions (even track 12 shows significantly faster RT than the visual-only) seem to show the several sources of confounding effects on RT. Therefore, the comparison of incongruent multimodal tracks and visual-only tracks did not support H1b that incongruent multimodal cue-target pairs will have longer RTs than those in the visual-only condition. Rather, all asynchronous conditions tended to show faster RT. This might be because sound's arousal effect increased drivers' attention level and thus, sped up the drivers' RT regardless of whether the sounds are related to the primary driving task or not [6]. The arousal effect might somehow cancel out the spatially incongruent cues' plausible delay effects. Overall, the data tend to support H1c as shown in Figure 3.

5.2. Temporal Rules (H2)

H2a and H2b are concerned with the temporal rules in crossmodal links. As hypothesized in H2a, the asynchronous multimodal pairs (Tracks 6, 7, 8, 9, 10 & 12) showed shorter RT than the visual-only baseline, either significantly (Tracks 6, 8, 9, & 12) or numerically (Track, 7 & 10). Therefore, H2a seems to be mostly supported by the results. The results support Posner's preparation function theorem [9] that priming auditory cues benefit reaction time. In H2b, we hypothesized that RT in the synchronous multimodal pairs would not be longer than that in the visual-only condition (based on crossmodal synesthesia). The majority of synchronous pairs (Tracks 2, 3, 4, 5, & 11) showed numerically longer RT than visual-only condition. This trend seems against H2b. Thus, results of RT did not support the synchrony benefit predicted by crossmodal synesthesia.

Why did crossmodal synesthesia not occur in this experiment? The Colavita bias [23] might be the reason. The Colavita visual dominance effect refers to the phenomenon where participants respond more often to the visual component of an audiovisual stimulus, by commonly neglecting the auditory component. In speeded audiovisual asynchrony discrimination tasks, Koppen and Spence investigated the influence of different Stimulus Onset Asynchrony. To many synchronous A-V pairs, the visual cue was actually perceived 12ms faster than the auditory cue which might lead to a prior-entry effect. In summary, generating auditory cues at the same time as visual cues might not result in the simultaneous processing necessary for crossmodal synesthesia.

5.3. Spatial-Semantic Conflict (H3)

In his spatial cuing task, Posner used only the non-verbal sound as auditory cues. The present experiment expanded the asynchrony benefit to the verbal cues. The addition of verbal cues created an interesting case: spatial and semantic conflict. The asynchronous (200 ms in this experiment) A-V pairs sped up response time either when there was no location-meaning conflict between A-V modalities (Tracks 8 & 12) or when the auditory cues were only spatially congruent with the visual target (Tracks 8 & 9). For the tracks having verbal cues, the spatially and semantically congruent groups had the shortest RTs among verbal pairs. (Track 3 had faster RT than Tracks 4, 5, & 11. Track 8 has faster RT than Tracks 9, 10, & 12).

H3a predicted that spatially incongruent and semantically congruent pairs would have longer RT. This was partly supported by Track 5. Track 5 showed the longest RT. Track 10 did not support this hypothesis, perhaps because its asynchrony improved RT. On the other hand, H3b predicted that spatially congruent and semantically incongruent pairs would have shorter RT. This was also partly supported by Track 9, which showed significantly faster RT than the visual-only. Track 4 did not support this hypothesis, perhaps because its synchrony degraded RT. Taken together, spatiality seems to be more powerful than semanticity in both cases (i.e., "where" information is more rapidly processed than "what" information is). However, the temporal dimension seems to have priority and confounds the results.

One interesting result came from spatially incongruent semantically incongruent pairs (Tracks 11 & 12). These had better performance than spatially incongruent and semantically congruent pairs (Tracks 5 & 10) because the spatial and semantic nature within the verbal cues were still consistent with each other despite being incongruent with the visual cue (e.g., visual cue directing the right, but auditory cue saying the word "LEFT" coming from the left speaker). The conflict within the verbal cue seems to have stronger effects than the conflict between A-V modalities.

5.4. Task Type and Speed-Accuracy Tradeoffs

Overall, the effects of the auditory cues on accuracy was very small. This could be explained by the distinction between visual scanning and visual tracking. Identifying the visual indication could be considered a visual scanning task. After changing the lane, keeping the lane position (by definition of PCL) could be the visual tracking task. As expected, auditory cues influenced the visual scanning task (reaction time) more than the visual tracking task (accuracy). However, there was also a trend of typical speed-accuracy tradeoffs. Most asynchronous auditory conditions improved reaction time, but most asynchronous auditory conditions seem to have lower accuracy than the synchronous auditory conditions. Triggering the response faster does not guarantee better or smoother control of the vehicle. Therefore, more research needs to be done to explore to what extent this trade-off could occur and whether it ultimately harms overall driving performance.

6. CONCLUSION AND FUTURE WORKS

We evaluated reaction time and accuracy of lane change test for verbal and nonverbal auditory cues manipulated along three dimensions (spatial, semantic, and temporal) in the presence of a visual cue. The results showed that the application of the multimodal displays could improve lane change test performance, but also showed that there are myriad interactions among variables.

Our results indicate that adding auditory cues could improve lane change test reaction time more than accuracy. The temporal dimension seems to be the most influential in performance. That is, preceding auditory cues improved reaction time. Spatially and semantically congruent auditory cues also facilitated reaction time. However, when these two dimensions conflict with each other, spatial congruency seems to have bigger impacts on performance. In other words, it is more difficult to ignore spatial location information than semantic verbal information just as in Barrow and Baldwin’s research [1]. Moreover, when there is conflict between auditory cues and visual cues, having consistency in auditory cues would be more important than inconsistency within the auditory cue and partial consistency with the visual cue. In-vehicle technology designers will want to consider the plausible trade-offs when designing the multimodal warning or alert system.

MRT suggests that well-designed multimodal interfaces can allow drivers to more efficiently process information in distinct channels. Furthermore, MRT can readily account for the results of the current experiment. However, MRT includes only verbal information processing regarding auditory modality. The empirical evidence of the present study using non-verbal auditory cues supports the necessity of updating the model [24]. Then, the model will be able to better explain and predict the effects of non-verbal auditory displays of the multimodal interfaces. The results also showed sound’s strong arousal effect, which can be better explained by the auditory preemption theory [25]. Certainly, more research is required to disentangle the various influences of auditory cues.

In future studies, it would be interesting to see the effects of the visual secondary task to increase driver workload. Given that Posner’s experiment using the 200 ms interval was not conducted in the driving domain, more asynchronous intervals can also be tested in the experiment to see if there is any different threshold in multimodal perception while driving. More research on the definition of a reaction timer will be helpful in the maneuver level driving task compared with the operational level (go/no-go) driving task. We also plan to conduct a similar study using a higher fidelity simulator, which provides a more realistic driving environment.

7. REFERENCES

[1] J. H. Barrow and C. L. Baldwin, “Verbal-spatial cue conflict: implications for the design of collision-avoidance warning systems,” in *Proceedings of the... international driving symposium on human factors in driver assessment, training and vehicle design*, 2009, vol. 5, pp. 405–411.

[2] C. D. Wickens, “Multiple resources and mental workload,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 50, no. 3, pp. 449–455, 2008.

[3] C. D. Wickens, S. J. Mountford, and W. Schreiner, “Multiple resources, task-hemispheric integrity, and individual differences in time-sharing,” *Hum. Factors*, vol. 23, no. 2, pp. 211–229, 1981.

[4] H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” *Nature*, vol. 264, no. 5588, pp. 746–748, Dec. 1976.

[5] W. Giang, E. Masnavi, and C. M. Burns, “Perceptions of Temporal Synchrony in Multimodal Displays,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2011, vol. 55, no. 1, pp. 1165–1169.

[6] C. Spence, “Crossmodal spatial attention,” *Ann. N. Y. Acad. Sci.*, vol. 1191, no. 1, pp. 182–200, 2010.

[7] R. W. Proctor, H. Z. Tan, K.-P. L. Vu, R. Gray, and C. Spence, “Implications of compatibility and cuing effects for multimodal interfaces,” in *Proceedings of the HCI International 2005*, 2005, vol. 11.

[8] J. D. Olsheski, “The role of synesthetic correspondence in intersensory binding: investigating an unrecognized confound in multimodal perception research.” Georgia Institute of Technology, 2014.

[9] M. I. Posner, R. Klein, J. Summers, and S. Buggie, “On the selection of signals,” *Mem. Cognit.*, vol. 1, no. 1, pp. 2–12, 1973.

[10] S. Sinnett, S. Soto-Faraco, and C. Spence, “The co-occurrence of multisensory competition and facilitation,” *Acta Psychol. (Amst).*, vol. 128, no. 1, pp. 153–161, 2008.

[11] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman, *Engineering psychology & human performance*. Psychology Press, 2015.

[12] C. Wickens, J. Prinet, S. Hutchins, N. Sarter, and A. Sebok, “Auditory-Visual Redundancy in Vehicle Control Interruptions Two Meta-analyses,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2011, vol. 55, no. 1, pp. 1155–1159.

[13] J. M. Pieters, “Ear asymmetry in an auditory spatial Stroop task as a function of handedness,” *Cortex*, vol. 17, no. 3, pp. 369–379, 1981.

[14] A. R. Mayer and D. S. Kosson, “The effects of auditory and visual linguistic distractors on target localization.,” *Neuropsychology*, vol. 18, no. 2, p. 248, 2004.

[15] J. H. Barrow and C. L. Baldwin, “Semantic versus Spatial Audio Cues: Is There a Downside to Semantic Cueing?,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2009, vol. 53, no. 17, pp. 1071–1075.

[16] J. H. Barrow and C. L. Baldwin, “Individual differences in verbal-spatial conflict in rapid spatial-orientation tasks,” *Hum. Factors J. Hum. Factors*

Ergon. Soc., p. 18720814553792, 2014.

[17] A. H. S. Chan and C. K. L. Or, “A comparison of semantic and spatial stimulus-response compatibility effects for human-machine interface design,” *Eur. J. Ind. Eng.*, vol. 6, no. 5, pp. 629–643, 2012.

[18] S. Mattes, “The lane-change-task as a tool for driver distraction evaluation,” *Qual. Work Prod. Enterp. Futur.*, vol. 2003, p. 57, 2003.

[19] ISO, “26022: 2010 Road vehicles–Ergonomic aspects of transport information and control systems–Simulated lane change test to assess in-vehicle secondary task demand,” *Int. Organ. Stand.*, 2010.

[20] H. Tattegrain, M.-P. Bruyas, and N. Karmann, “Comparison between adaptive and basic model metrics in lane change test to assess in-vehicle secondary task demand,” in *21st esv conference*, 2009.

[21] T. D. Wickens and G. Keppel, “Design and analysis: A researcher’s handbook.” Englewood Cliffs, NJ: Prentice-Hall, 2004.

[22] C. Spence and J. Driver, *Crossmodal space and crossmodal attention*. Oxford University Press, 2004.

[23] C. Koppen and C. Spence, “Audiovisual asynchrony modulates the Colavita visual dominance effect,” *Brain Res.*, vol. 1186, pp. 224–232, 2007.

[24] M. Jeon, “How Is Nonverbal Auditory Information Processed? Revisiting Existing Models and Proposing a Preliminary Model,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2016, vol. 60, no. 1, pp. 1529–1533.

[25] C. D. Wickens, S. R. Dixon, and B. Seppelt, “Auditory preemption versus multiple resources: Who wins in interruption management?,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2005, vol. 49, no. 3, pp. 463–466.