2016

MULTISENSORY CUE CONGRUENCY IN LANE CHANGE TEST

Yuanjing Sun
Michigan Technological University, ysun4@mtu.edu

MULTISENSORY CUE CONGRUENCY IN LANE CHANGE TEST

By

Yuanjing Sun

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2016

This thesis has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Applied Cognitive Science and Human Factors.

Department of Cognitive and Learning Sciences

Thesis Advisor: *Dr. Myounghoon Jeon*

Committee Member: *Dr. Shane T. Mueller*

Committee Member: *Dr. Andreas Riener*

Department Chair: *Dr. Susan Amato-Henderson*

Table of Contents

ABSTRACT

Abundant information due to mobile internet and technology has even extended to the driver's seat. Now, it is common to see a driver interacting with multiple systems while driving. Multimodal in-vehicle technologies intend to facilitate multitasking while driving because the synergy from multimodality is able to reduce cognitive effort in processing information. The present study aims to investigate how congruent or incongruent (temporal, spatial, & semantic) multimodal cues (auditory & visual) facilitate or impair driving performance.

Twenty-six young drivers participated in the Auditory-Spatial Stroop experiment in a lane change scenario with different combinations of multimodal signals. The response time, accuracy, and mean deviation in lane change test were separately analyzed.

As a result, the asynchronous (i.e., with advanced auditory cues) congruent cues seemed to enhance reaction time over visual-only signals in the lane change test. However, when the spatial congruency and the semantic congruency conflicted with each other, the spatial congruency had a stronger effect than the semantic congruency. There was not much improvement in accuracy with auditory cues. However, with accuracy, only spatially incongruent verbal cue negatively impacted performance compared to the visual-only condition. Mean deviation analysis did not show any clear results. The results of the present thesis can be applied to the design of the entire in-vehicle auditory scene. However, a cautious approach is required to avoid location-meaning conflict in auditory signals, which will lead to cognitive overload and plausible risk.

1. INTRODUCTION

The mobile internet has extended interactions with web applications (e.g., Google Maps, Facebook, Yelp, etc.) to a driver's seat. Multitasking while driving becomes an inevitable challenge not only to drivers, but also the whole automobile industry. Multiple Resource Theory (MRT) (2008) suggests that in-vehicle technologies convey information through other modalities than vision. In fact, speech recognition or vibrotactile notifications are pervasive in vehicles nowadays. In spite of the fact that a multimodal interface allows drivers to process more information in parallel, it still occupies part of attentional resources. Does more information always mean more facilitation? With bad design, multimodal displays might cause information overload or even degrade performance, which will lead to a safety hazard on the road.

The blooming of multimodal interfaces has often occurred regardless of the limitation of human multisensory information-processing (Ho & Spence, 2012). 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, it might not always benefit multiple tasks, especially when modalities conflict with one another at the same time. Previous studies (Proctor & Van Zandt, 2008) categorize the multimodal signaling issue into three conditions: (1) Whether there is strong automation between signal-response, (2) Whether the multimodal signal causes conflicted spatial attention, and (3) Whether the driver can schedule responses according to the appropriate emergency level. In multitasking theory, three main factors are available resources, reserve capacity, or schedule issue.

The present study aims to compare multimodal cues with unimodal (visual-only) signaling in a driving scenario. In particular, auditory cues could improve driver responses to a visual target in a seriously controlled driving environment. 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.

The hierarchical model of driving behavior

Michon (1985) developed a hierarchical model of driving behavior that includes three levels: operational, maneuvering, and strategic. The lowest operational level involves immediate vehicle controls, such as braking, shifting, etc. Reactions at the operational level are regarded as a single task (two-choice detection task). Fisk showed that brake or steering wheel control was regarded as automatic reflexes, regardless of variability in the driving scenarios (as cited in Trbovich, 2006). The maneuvering level involves negotiation of common driving situations (e.g., negotiating curves, intersections, gap acceptance in overtaking or entering the traffic stream, performing lane change maneuvers, and obstacle avoidance). Therefore, actions at the maneuvering level require perceptually processed signals and integration of multiple pieces of visual and spatial information in the driving environment. The strategic level involves general trip planning, including setting trip goals (e.g., minimizing time, avoiding traffic), selecting routes, and evaluating the cost and risk associated with alternative trips. The present proposal focuses on the maneuvering level of Michon's hierarchy because actions at the maneuvering level appear to be the most susceptible to interference between perceptual and cognitive levels. The impact of various combinations of multimodal representation in design of IVIS will be directly reflected in the maneuvering level performance.

Lane change test (LCT)

Given the complexity and multifaceted nature of driving, the prototype of a multisensory interface needs to be tested before implementation. There is always a question about the validity of laboratory experiments towards real cases. It is challenging to simulate multitasking in driving in laboratory studies because splitting a continuous process into separate parts, which are identical to theory framework, is difficult.

Of course, there have been some attempts. Several industry standards have regulated the safety documents about how to integrate IVIS (e.g., telematics in car networks such as PND or intelligent transportation systems (ITS), various infotainment systems, or consumer electronic devices) with dashboard controls. An European project, "Adaptive Integrated Driver-vehicle interface" (AIDE) (Engström et al., 2004), and a U.S. project, "SAfety VEhicle using adaptive Interface Technology" (SAVE-IT) (Zylstra, Tsimhoni, Green, & Mayer, 2003) implemented evaluations of potential distraction risks caused by IVIS. They made efforts on constructing methodologies and conducting empirical studies to measure driving distraction. However, both of these two projects used an interruptive secondary task and measured performance degradation as metrics of distraction. However, distraction is a dynamic variable that continuously impacts on driving performance. Also, independent driving simulators and scenario settings make it difficult to compare similar studies.

To compensate for these issues, a relatively new approach in examining the distraction potential of IVIS was introduced. The Lane Change Test (Mattes, 2003) was devised to evaluate driving performance while drivers interact with in-vehicle assistive devices. The Lane Change Task (LCT) was developed within the Project ADAM (DiamlerChrysler, BMW), which is a simple laboratory dynamic dual-task method that quantitatively measures performance degradation in a primary driving task. In real-world situations, driving a vehicle requires a driver to perform several tasks based on variations in the environment. In this situation, several stimuli often require driver responses in a rapid succession (Proctor & Van Zandt, 2008). The dual task paradigm aims to test how well a person can select and coordinate multiple responses (e.g., driving as a primary task and menu navigating on telematics system as a secondary task). On the other hand, a leading car–following test only requests a participant to keep distance by using a brake pedal. This type of experiment is only categorized into a simple reaction task because the driver only needs to have a go/no-go choice, which is largely related to automatized reflex, instead of a task with speed-accuracy trade-off. The current study tried to test multimodal (audios and visuals) interfaces through a revised LCT test in which participants will have four options (left, leftmost, right, and rightmost) that probe the maneuvering level in Michon's hierarchy. Therefore, it belongs neither to the dual task nor the traditional go/no-go task.

The dual task paradigm usually included several signal inputs, which makes it difficult to identify which source causes the benefit or degradation. The present thesis tries to control the input of dual task and focus on the direct causal-effect mechanism between multimodal cues and driving performance. Thus, I chose an auditory location memory task as the secondary task (i.e., Count how many auditory cues heard from either left or right side and report the total number in the end of each trail.). It simulated a continuous distraction but without extra signal inputs.

In terms of performance output, LCT uses both event detection and maneuver execution as two variables and thus, it is respectively sensitive to responses caused by different types of perception processing (i.e., bottom-up and top-down). The overall mean deviation in lane change path measure encompasses two tasks; (1) a lane change initiation task: responding to road signs or auditory commands and (2) a maneuvering task: maneuvering quickly and efficiently into a given lane and maintaining the lane position between two consecutive signs. The variance in these different parameters can help identify which theory or model can explain the appearance or absence of multimodal facilitation under specific primary and secondary task (controlled task type and perceptual load). Thus, LCT would be a suitable driving test to compare the different impact between nonverbal (spatial) and verbal (spatial and semantic) cues on driving performance.

2. LITERATURE REVIEW

The information processing framework can be divided into various sub-processes, depending on different perspectives. Both multimodal/cross-modal facilitation and inhibition have been studied with different mechanisms and theories. I reviewed models and theories that are closely related to this proposal: Multiple resource theory (Wickens, Hollands, Banbury, & Parasuraman, 2013); spatial rule (as cited in Spence, 2010) and temporal rule; type and demand of visual tasks (as cited in Wickens, Prinet, Hutchins, Sarter, & Sebok, 2011); and Colavita bias (Colavita, 1974). However, note that those are selected ones and not intended to be exhaustive.

2.1. Cross-modal interference, facilitation, and Multiple Resource Theory (MRT)

Wickens' multiple resource theory is to predict interference between concurrently perceived signals. It is composed of four dimensions (Figure 1), which includes stages, modalities, accesses ("codes" in the earlier version) and responses. The MRT suggests that two tasks demanding separate resources along these four dichotomous dimensions can improve the overall time-sharing performance and impair individual tasks less than tasks occupying the same resources. For the "modality" dimension, it suggests that people have independent sub-perceptual channels (i.e., auditory vs. visual) to extract signals from environments. For the "access" dimension, verbal and spatial resources are respectively stored with different "codes". For the "stage" dimension, resource for perception is separated from resource for responding.

MRT provides the theoretical background that using multimodal information presentation might minimize distraction effects on driving performance. Conversely, it is challenged by multisensory illusions, such as McGurk illusion and Ventriloquism illusion. McGurk illusion (McGurk & MacDonald, 1976) describes a phenomenon that a sound of /ba/ tends to be perceived as /da/ when it is paired with a visual lip movement /da/. This contrary example of perception illusion is raised by simultaneous, but incongruent audio-visual processing. According to Wickens' MRT, as long as information is coded in verbal, it should be independently perceived without the interference of visual spatial information. However, MRT does not provide an explanation about how spatial *auditory* information (whether verbal or nonverbal) is processed.

Also, multimodal cues could be beneficial in one phase but have reversed effects in the other phases. Liu and Jhuang (2012) investigated the effectiveness of an IVIS on emergent response and decision making performance. They used touch screen to record participants' RT respectively from four stages: detection, location, identification, and decision making according to sub-steps of the task. First, the pre-alert cue for the visual warning information (presented twice each at 1 Hz frequency) appeared and followed by the auditory warning information, a tone "Dong". Then participants were asked to point the location and made a multiple choose of the warning content in touch screen. The intervals between signal onset and separate touches were collected as Reaction Time (RT) and accuracy data in four phases. Spatially congruent auditory cues only improved RT in the detection and location stage, but redundant displays prolonged RT in the decision making stage.

It has been well established that driving requires acquisition of visual and spatial information. PNDs provide verbal directions in driving, while lateral collision avoidance systems send out a directional alert to a potential hazard. This type of situation might constitute a “spatial Stroop paradigm”. The original color-word Stroop task investigates how incongruent stimuli (e.g., color of text versus meaning of text) would influence responses (Virzi & Egeth, 1985). In the spatial Stroop paradigm, the stimuli are related to both spatial and semantic properties. For example, the word, “LEFT” or “RIGHT” is presented in a corresponding or opposite position from its meaning. In this line, Baldwin (2012) used an “Auditory-Spatial Stroop paradigm” to examine which one (between spatial cue and semantic cue) is more influential in incongruent cue combination. The proposed study used an Auditory-Spatial Stroop experiment to examine how spatially or semantically incongruent audio-visual cues would impact driving performance.

Wickens’ Multiple Resource Theory (MRT) (Wickens, Mountford, & Schreiner, 1981) has served to predict or analyze interference between concurrently perceived signals. The MRT suggests that two tasks demanding separated resources can improve the overall time-sharing performance. 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 Ventriloquism illusion (McGurk & MacDonald, 1976). The conflict between MRT and multisensory illusion leads to a further step of looking into the rules of how multisensory perception influences information processing.

2.2 Regularities in Cross-modal Facilitation

The conflict between MRT and multisensory illusion leads to a further step of looking into the rules of how multisensory perception influences information processing. Some researchers have suggested that responses to simultaneous multisensory stimuli can be faster than responses to the same stimuli presented in isolation (Spence & Driver, 2004). Cross-modal synesthesia describes a condition in which a person experiences sensation in one modality when a second modality is stimulated (Olsheski, 2014). The stimulation of one sense elicits an additional experience transduced from other sensory channels. For example, a form of synesthetic association example includes the relationship between auditory pitch and visual size, where lower frequency tones are associated with large objects and higher frequency tones with small objects.

2.2.1. Cross-modal facilitation versus inhibition in spatial and temporal aspects

The degree of multimodal benefits follows both (1) spatial rules and (2) temporal rules. However, several conflicting studies make it difficult to identify exactly where the facilitation derives from.

Spatial Rules

In Spence's (2010) review of cross-modal spatial attention, the *mean spatial cuing effect* was defined as the RT performance benefit on ipsilaterally (i.e., the cue comes from the same direction as the target) cued trials over contralateral (i.e., the cue comes from the opposite direction to the target) cued trials for each cue type. A possible mechanism 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 one's spatial attention to the cue direction. Another experiment also supports such an explanation. 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 the spatially non-corresponding mapping (Proctor, Tan, Vu, Gray, & Spence, 2005).

Temporal rules

A temporal rule outlines that responses to multimodal cues would benefit from perceived synchrony of the multimodal inputs thanks to maximal overlap in respective periods of peak activity. However, the synchrony benefits may not explain every case. Posner, Klein, Summers, and Buggie (1973) asked observers to respond with a left or right key to a target that occurred to the left or right of a vertical line. On each trial, the time interval between priming warnings and the target signals varied from 0 msec to 400 msec. Participant's response times were plotted like a U-shape with the time interval increased. In other words, the RT became slower at the points when priming warning is either 0 or 400 msec than RT under 200-msec-preceding warning. The U-shape plot was also denoted as "preparation function" to describe how the response time appeared as a function of the SOA between the priming warning and the target stimulus. Two hundred msec was the bottom of this U-shape function, at which point, the quickest response was recorded. Therefore, the present study selected 200 msec as a preceding timing as the asynchrony condition in contrast with the synchrony condition.

2.3. Type and Demand of Visual Tasks for Cross-modal Facilitation

Although multimodal interfaces are often considered better in time-sharing performance (Wickens et al., 2011), multimodality does not always win over unimodality. Sinnett, Soto-Faraco, and Spence (2008) manipulated perceptual load (frequency of visual targets) and working memory load (alternative numbers of response) to compare the redundant gain under these two experimental settings. The result indicated that both multisensory facilitation and inhibition can be demonstrated by changing the task type and visual demand. They found an explanation from Broadbent's study. Broadbent claimed that perception has a limited capacity in early-selection of attention (Broadbent, 1958) but processes all of the available stimuli in an automatic and mandatory fashion as suggested by late-selection theorists (e.g., Deutsch & Deutsch, 1963) until the free capacity is drained out (as cited in Santangelo & Spence, 2008). The perception load changed when Sinnett et al. manipulated the frequency of visual target. To some point, the perception capacity is drained out and then turns the multimodal perception as a burden because distractors come with the redundancy. Lavie's hybrid model (Lavie, Hirst, De Fockert, & Viding, 2004) agreed with the early/late attention selection hypothesis that excluding distractors depends on the availability of free perceptual resource. In sum, the control of perceptual load of a given task in the experiment is important in a multimodal test. However, perceptual load is difficult to be measured or be compared between respectively conducted experiments.

Multimodal benefits do not always produce better time-sharing because of the disruption from auditory cues in a visual-visual (Vv) tracking task. In A-V redundancy studies, ongoing tasks (OT), usually termed as tasks, require continuous visual attention. In the context of OT, there are periodic "interrupting tasks" (IT) that are discrete in nature. To clarify the term, the capitalized "V" means visual OT whereas, lower case "a" or "v" indicates the modality of the interruptive tasks for tasks mentioned in Wickens and his colleagues' mata-analysis (Wickens et al., 2013).

Wickens et al. (2013) suggested that a redundant display may benefit only to a visual scanning task but not to the ongoing visual tracking task. A meta-analysis of 29 studies comparing visual-auditory (Va) tasks with visual-visual (Vv) tasks has shown that Va has a discrete intermission and Vv was a relatively continuous task. Using auditory presentation of a discrete task resulted in a significant 15 percent advantage over visual presentation. The auditory advantage enlarged when the two visual inputs were end-to-end (Wickens, Prinet, Hutchins, Sarter, & Sebok, 2011). In other words, the auditory cues are more helpful when the interval between two visual inputs is short (i.e., visual perceptual load is high). It can be inferred that the A-V facilitation would be more likely to occur in visually-demanding tasks. Also, in terms of task type, the A-V facilitation would be more likely to occur when the secondary task is visual scanning task than visual tracking task.

2.4. How Colavita visual dominance effect could impact SOA selection scope in a sensation level

Colavita visual dominance effect

Stimuli intensity (e.g., brightness, audibility, etc.) impacts how well a stimulus can be identified in a sensation level. The Colavita visual dominance effect (Colavita, 1974) refers to the phenomenon where participants respond more often to the visual component of an audiovisual stimulus, when audiovisual stimuli were presented concurrently. Theorists have proposed that the Colavita effect demonstrates a bias toward visual sensory information because the presence of auditory stimuli is commonly neglected during audiovisual events. Koppen and Spence (2007) conducted a series of Temporal Order Judgment experiments to determine the Point of Subjective Simultaneity. Their findings helped to construct a scope of SOA for speeded discrimination task paradigm because the sensation of audiovisual asynchrony not only depends on physically temporal difference but also on the sensation channels of human beings. The temporal window for audiovisual integration (humans who do not suffer from sensory difficulties perceive audio and visual signal at the same time) was recalibrated -65 msec to 89 msec (negative means auditory stimulus precedes over visual stimulus).

The temporal window need to be noticed for researchers in selection of appropriate SOAs for sake of different purposes. To ensure the participant perceive the speech priming asynchrony, the speech cue should precede at least 85 msec ahead of any visual stimulus according to another following study (Vatakis, Navarra, Soto-Faraco, & Spence, 2007).

2.5. The Current Study and Hypotheses

To assist driving, IVIS designers have tried to represent information through multimodal channels. However, respective multimodal displays might cause information overload, and thus, impairing drivers' reaction time and driving performance. Understanding different mechanisms involved in multisensory perception is important to choose appropriate modalities for display. The proposed study intends to ascertain the decisive mechanism(s) in multisensory perception. By reviewing various models and empirical studies, I have found that competing rules and results are involved in the explanation of either the facilitation or inhibition of multimodal cues. Key metrics of the effectiveness of multimodal interfaces consist of three aspects: 1) Whether the salience of the two stimuli presented in different modalities are large enough to avoid the Colavita effect; 2) Whether the degree of SOA is controlled within a proper span that the multimodal signals are processed to facilitate performance results; and 3) Whether specific incongruency (e.g., spatial, semantic, or temporal) is superior to other congruent gain under a controlled perceptual load task. To ensure safe use of multimodal IVIS in driving, any multimodal interface should be tested in the context of the presence of other IVIS. That is, an ecological environment can maximize the effectiveness of using multimodal IVIS for road safety.

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, I constructed three major sets of hypotheses:

Hypothesis 1 concerns the spatial rules: Spatially congruent A-V pairs will have shorter reaction time (RT) than the visual-only condition (H1a). Spatially incongruent A-V pairs will have longer RT than the visual-only condition (H1b). 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 (H1c).

Hypothesis 2 concerns the temporal rules: Asynchronous (i.e., preceding auditory cues) A-V pairs will have shorter RT than the visual-only condition (H2a). Synchronous A-V pairs will not have longer RT than the visual-only condition (H2b). The ‘‘preparation function’’ is the response time as a function of the SOA between the priming warning tone and the target stimulus. Two hundred msec SOA is the bottom of the U-shape plot of the preparation function.

A verbal cue cannot be simply categorized as auditory modality because it has two dimensional properties, including both semantic congruency and spatial congruency. It would be important to see whether the spatial congruency has a larger impact on RT than the semantic congruency if the verbal cue would be sent out from the single channel or one side of the driver.

Hypothesis 3 concerns spatiality-semanticity conflict. Hypothesis 3A: When verbal cues have spatiality-semanticity conflict, the conflict pairs (SpCSemIc or SpIcSemC) will have longer RT than consistent pairs. Hypothesis 3B: When verbal cues are spatially incongruent but semantically congruent with visual targets, RT will still be slower than the visual-only condition. Barrow and Baldwin (2009) showed that it is more difficult to ignore spatial location information than semantic verbal information when the two pieces of information conflict with each other.

3. METHODS
3.1. Participants

Twenty-six participants (23 males, 3 females; M_{Age} = 20.6, SD_{Age} = 2.3; $M_{YearOfDriving}$ = 4.5, $SD_{YearOfDriving}$ = 2.86) were recruited from the Michigan Technological University undergraduate population via the SONA System, web-based recruitment software. Participants were expected to be above 18-year-old English native speakers. Each participant had a valid driver's license and at least 2 years from the issued date.

3.2. Experiment design and stimuli

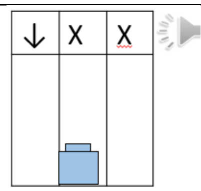
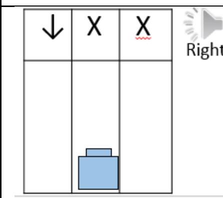
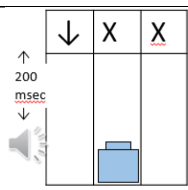
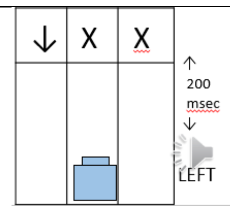
3.2.1. Experiment Design

The experiment is a within-subjects design. Each participant performed a total of fourteen tracks. Two of them were visual-only tracks which served as a baseline. In the twelve multimodal tracks, four were nonverbal tracks and eight were verbal tracks.

For nonverbal tracks, there were two dimensions in time wise and spatial wise. Since a visual target appeared in every track, it was the reference when using "congruent" or "incongruent" to name the group. For example, the abbreviation "SpC" means **sp**atially **c**ongruent to the visual target, whereas "SpIc" means **sp**atially **i**n**c**ongruent to the visual target. In the timing wise, I have **syn**chrony vs. **asyn**chrony conditions ("Syn" vs. "Asyn"), indicating the temporal gap of audio cues towards visual target. The stimulus onset asynchrony (SOA) (i.e., A-V asynchrony) between audio-visual stimuli was determined based on previous studies (Chan & Or, 2012; Proctor & Vu, 2006; Santangelo & Spence, 2008), which showed the optimal performance.

For the verbal cues, it had the third semantic dimension which makes eight combinations (2 spatial * 2 timing * 2 semantic). Some examples in Table 1 indicate different types of A-V combinations encountered by participants.

TABLE 1. ILLUSTRATION OF AUDIO-VISUAL COMBINATIONS OF MULTIMODAL CONDITIONS

	SpC	SpIc	SpIcSemC	SpIcSemIc
Syn				
Asyn				

3.2.2. Visual Stimuli

Each track had 18 lane change signs as well as one “START” sign in the beginning and one “FINISH” sign at the end. The "Lane Change" signs appeared in an overhead position of a gate on the simulated roadway. They were composed of one check mark and two crosses in three separate black borders (A-V interaction in Table 1 and snapshot in Figure 1). The borders were two-meter width and one-meter height as listed in ISO26022:2010. The signs were programmed to appear when the car reached 40 meters ahead of the gate. It was programmed to last on the screen for as 350 msec long as the auditory cue last.

FIGURE 1. THE OVERVIEW OF EXPERIMENT SETTINGS AND DRIVING SCENARIOS

3.2.3. Auditory stimuli (nonverbal and verbal cues)

Four non-verbal stimuli and four verbal stimuli were used as auditory cues in twelve tracks out of fourteen in total. The four verbal cues were “LEFT” “RIGHT”, “LEF-LEFT” and “RIGH-RIGHT”. The nonverbal cues were normalized as equal duration of 350msec at the volume level of 60 dB. The length and loudness of auditory cues were produced by reference to similar demands of the perceptual-motor experiments conducted by previous researches (Chan & Or, 2012). They examined the effect of semantic and spatial stimulus-response compatibility by using auditory cues in a single response task. Considering the similarity of task complexity and overall response window (4.1 sec), all auditory stimuli in this study 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 (Fromtexttospeech.com, 2015) at medium speed with a female voice (Laura, US English).

Sped-up verbal clips, “LEF-LEFT” and “RIGH-RIGHT” indicate the direction of double-lane change (i.e., from left most lane to right most lane or vice versa). For example, I imported the original TTS speech “LEFT” to Audacity 2.1.0 version and replicated the word “LEFT” to two audio tracks. For the first audio track, the first vowel “LE” was remained. For the second audio track, the full word “LEFT” was remained. The last step was to combine the two audio tracks and shrink the duration time of “LEF-LEFT” to the 350 msec. Audacity has “change tempo” effect to adjust the length of audio clip without changing the pitch.

Verbal cues had two levels of congruency: 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, when the visual cue indicates change to the left lane, the participant would hear a verbal cue, “LEFT” coming from the right speaker. This situation counts as semantically congruent and spatially incongruent condition.

3.3. Scenario

The simulated track length was 3,000 m, corresponding to around 1.5 minutes of driving at a constant 110 kph (70 mph) for each track. The 18 lane change signs were spaced approximately within 150 meters, corresponding to a lane change maneuver every

4.1 seconds. This scenario setting is different from the ISO26022:2011 because of the low secondary task demand (auditory memory task). The ISO26022 standard did not specify secondary task choice. The standard secondary task (e.g., SuRT v.2.1) (Young, Lenné, & Williamson, 2011) has much higher task demand mostly coming from visual gaze distraction. However, the goal of current study is to compare three (spatial, semantic, and temporal) properties between verbal and nonverbal cues under a discrete visual task. Auditory display has cognitive distraction other than visual distraction. That is why I increased the constant speed from the 60 kph (ISO standard recommend) to 110 kph. The sped-up scenario increased the perceptual workload by increasing the frequency of event and shrank the response window as well. In addition, it constrained the whole experiment time within an hour to avoid fatigue caused by overtime experiment.

3.4. Apparatus

A low-fidelity driving simulator based on OpenDS v2.5 (http://www.opends.eu/) was implemented to perform the Lane Change Test (LCT) (Mattes, 2003). The LCT was developed within the Project ADAM (DaimlerChrysler, BMW), which is a simple laboratory dynamic dual task method, which quantitatively measures performance degradation in a primary driving task. The primary task requires a participant to drive in a simulated straight three-lane road containing a series of lane changes defined by signs.

Visual cues were displayed on a 39″ Samsung TV with a refresh rate of 60 Hz. A screen (brightness 300 cd/m^2, contrast ratio 500:1, minimum resolution of 1920*1080 pixel was for driving scene released. The horizontal viewing angle to the display for the road scenery (monitor or screen) was between 20 degree and 55 degree. The eye-to-display distance was no less than 60 cm. The horizon of the visual scene was between – 5 degree and +5 degree from the participant's eye point height (ISO, 2010). The audio cues were played through JVC HA-RX300 stereo headphone. The simulated vehicle position was controlled by a Logitech G27 wheel/pedals combinations.

The Auditory-Spatial Stroop experiment was developed on the basis of the embedded the ReactionTest scenario in OpenDS 2.5. I re-implemented the LCT Toolkit in OpenDS and made some modifications according to ISO26022-2010. Researchers can manipulate the timing and multimodal combinations of lane-change signs to capture different driving patterns under different conditions.

3.5. Procedure

After reading and signing the consent form procedure, a participant was given a video clip about an overview of the experiment and how to use a driving simulator. Before the experiment started, the experimenter helped the participant adjust the sitting position in the driving simulator to make sure that each participant drives in a comfortable condition. Then, a video clip gave instructions about how to quickly and efficiently change lanes when the lane change symbol appears in a training run. The video clip allowed all participants to get the same instructions on how to conduct a lane change test. Then, participants completed a training track, containing all possible combinations of multimodal signals which would appear in the following driving task. Also, an equivalent hearing test was given to the participant in the training trial. The participants repeated a standard list of words (LEFT, LEF-LEFT, RIGHT, RIGH-RIGHT) given through the headphone at various levels of loudness. 50% correctness was a pass for that test. The training track used gradient loudness audio file. A RT histogram popped out when the participant finished the training track. The real experiment started when the participant confirmed he or she understood the whole process. To ensure the participants actually hear the auditory stimuli during test, they were required to report the total number of the audio cues they heard either from the left side or right side in the end of each track. The experimenter recorded their counting accuracy as performance of secondary task. Without inserting extra input, the auditory counting task served as a cognitive secondary task to increase the cognitive workload. Experimenter asked for a total number of auditory cues coming from either left or right side after each trail.

3.6. Design and Conditions

Table 2 is the experimental design of twelve multimodal conditions. “78%” indicated that five out of eighteen signs were distractors in the track. For example, in SynSpC track, participants were supposed to hear 13 auditory cues coming from the same side of the visual target, but five auditory cues coming from the opposite side. Such arrangement could eliminate participants to get familiar with the pattern of each track.

TABLE 2. EXPERIMENT DESIGN AND AUDIO-VISUAL MAPPINGS IN SPATIAL SEMANTIC AND TEMPORAL DIMENSIONS.

	Nonverbal Cue		Verbal Cue			
	78% SpC	78% SpIc	78% SpCSemC	78% SpCSemIc	78% SpIcSemC	78% SpIcSemIc
Sync	Track 1	Track 2	Track 3	Track 4	Track 5	Track11
Asyn	Track 6	Track 7	Track 8	Track 9	Track 10	Track12

Counterbalancing the track order

The order of 14 tracks was counterbalanced as shown in Table 3. Participants were randomly distributed into four groups. Order 1 & 2 were reversed sequential orders. Order 3 split the tracks in the middle to the two extremes. In this way, the order effects can be minimized. To reduce participants' adaptation to repeated patterns, asynchrony, congruency and modality in each order were considered.

TABLE 3. FOUR ORDERS OF EXPOSURE SEQUENCE OF FOURTEEN TRACKS

	Start--->			**End**
Order 1	Track0	Track8->2->11->6->4->7->1->10	Track13	5->12->-3->9
Order 2	Track0	Track9->3->12->5->10->1->7->4	Track13	6->11->-2->8
Order 3	Track0	Track7->4->6->12->2->8->9->3	Track13	5->11->10->1
Order 4	Track0	Track1->10->11->5->3->9->8->2	Track13	12->6->4->7

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 was inserted within the 7th to 10th run to see the trend of the learning effect. It is a method to evaluate how the learning effect of LCT would interfere with the dependent variables (RT and lane deviation) with time (Petzoldt, Brüggemann, & Krems, 2014).

3.7. Evaluation criteria and Metrics

3.7.1. Lateral control reflects workload of cognitive task

Engström and Markkula (2007) have examined the sensitivity of two new LCT metrics: a path control (high-pass filtered standard deviation of lateral position; SDLP) and sign detection/recognition (Percent correct lane; PCL) to distinguish visual and cognitive tasks. Path control performance was quantified by means of the high-pass filtered (at 0.1 Hz) standard deviation of lateral position (SDLP), calculated for an entire track, where the lateral position was measured relative to the road (and not relative to a specific lane). The purpose of high-pass filtering was to remove the low-frequency effect of the lane changes. Results revealed that the two types of distraction each impaired LCT performance differently. The visual, but not cognitive, tasks led to reduced path control, while the cognitive, but not visual, tasks affected detection and sign recognition and responses.

Lateral control and event detection were found in different levels of sensitivity in the evaluation of task demand in LCT (Young, Lenné, & Williamson, 2011). Lateral control metrics were found to be sensitive to detect different workload of cognitive tasks, while event detection metrics were less able to discriminate different task demands (Young et al., 2011).

3.7.2. Initiation delay reflects event detection

Mean delay in lane change initiation was defined as the time (in seconds) elapsed between the moment the sign appears (40 meters before the sign reached) and the initiation of the lane change. The metric is only applied to correct lane changes, as determined by the method described in the previous section. The initial point was defined in terms of the most significant steering action towards the new lane, which was identified by means of the following method, composed of three steps (ISO, 2010).

3.7.3. System log data and adaptive calculation of mean deviation of trajectory

Reaction Time and PCL were two direct metrics of performance. The car position parameters (i.e., positional coordinates and steer angle) were automatically recorded by the driving simulator at the sampling rate of at least 10 Hz, ISO3.3.5. The reaction to the stimulus is measured as the time span between stimulus and a steering wheel angle outside of the ordinary lane keeping range. The Reaction Timer is activated simultaneously when the earlier cues appear. It can output the milliseconds taken when the car maintains straight in the targeted lane for 800 msec (This 800 msec would be subtracted from the RT outputs). The maximum RT window for correct completion lane change is 4.1 seconds after the lane change sign, which has been defaulted in OpenDS Reaction Task settings. Such a setting excludes overshooting from recording correct lane-change maneuver. OpenDS has a built-in measurement engine that can be configured to trigger specific measurements (e.g., reaction time), checking the validity of measures, and finally storing data of interest in txt-type log files.

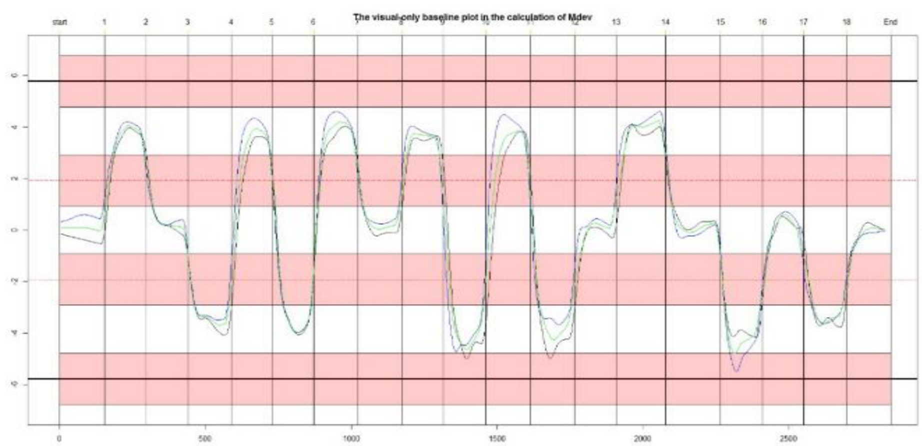


FIGURE 2. ROAD AREA PARTITION TO CATEGORIZE THE CORRECTNESS OF LANE CHANGE ALTERED FROM THE GUIDELINE (TATTEGRAIN, BRUYAS, & KARMANN, 2009) PUBLISHED IN NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION WEBSITE. THE GREEN TRAJECTORY IS THE BASELINE IN THE CALCULATION OF MEAN DEVIATION.

3.7.4. Percent of correct lane PCL (accuracy of lane-change)

The accuracy of lane-change completeness is quantified in terms of the percent correct lane (PCL). For each track, the Percent Correct Lane-change (PCL) was measured as the fraction of the consistent lane choices that were correct. Figure 2 is a diagram to show how the system distinguish the “correct lane-change” from “erroneous lane-change”.

The diagram was recreated according to the guideline (Tattegrain, Bruyas, & Karmann, 2009) in National Highway Traffic Safety Administration website (copyright permission in Appendix. B). The measurement in their study divided the three-lane-road into different zones. The white area in Figure 2 corresponds to a correct position in lane1 (left lane), lane2 (center lane) or lane3 (right lane), while the pink zones corresponds to out of valid positions. As long as the lateral position of the car maintains 75% of its trajectory within the valid area between two signs, the response is a “Correct LC”. Otherwise, the reaction timer outputted an “NA” instead of RT for being in invalid area. The correctness of each lane change can be categorized as (1) “Correct LC”: the end position of the driver is in the attended lane; (2) “No LC”: the driver is in the white zone at the same lane from the start till the end positions; and (3) “Erroneous LC”: the end position of the driver is in valid area but the portion trajectory in pink area is big enough to cause a hazard.

The lane change sign being displayed at a distance of 40 meters before the sign position. In this way, participants have 110 m to complete lane change and maintain straight in the target lane. Lane keeping maneuver distinguishes two successive lane-change maneuvers and provides a buffer if participants have erroneous lane-change in the previous sign. The segmented distance from the last sign will not influence the start position of the upcoming sign.

3.7.5. Lane deviation calculation

Mean deviation (Mdev) comes from the total intersection area between baseline and the driven course in each condition. The baseline is the average trajectory from two visual-only tracks as the green line in Figure 2. With the Mdev, I can compare the lane-change behavior between the baseline run and the condition run. In addition, I can obtain the individual differences by comparing every participant's baseline run with the optimal curve.

4. RESULTS

Data were collected from 26 participants in four orders (Table 3). (Order 1 had 9 participants. Order 2 had 8. Order 3 had 4. Order 4 had 5.) Although there is a violation of the equal variance assumption on RT by gender, the violin plots of RT and Accuracy distribution by gender show that female's average RT and accuracy were within the scope of the 1st quantile and 3rd quantile of male. In other words, the unbalanced gender distribution did not skew the mean of all participants.

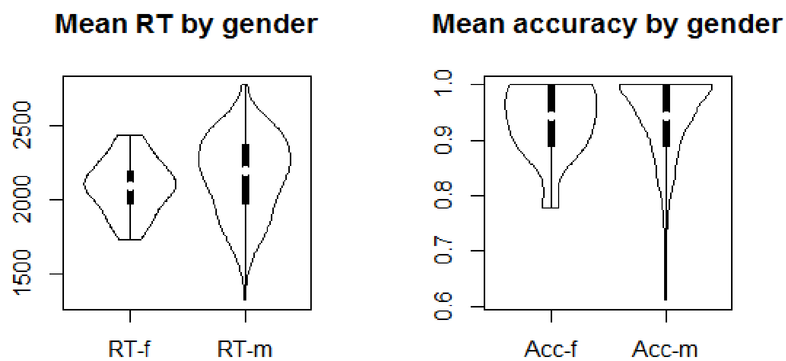


FIGURE 3. VIOLIN PLOT OF RT AND ACCURACY OF LANE-CHANGE BY GENDER. NOTES: THE WHITE SPOT IS MEAN OF THE GROUP AND THE BLACK BAR IS THE QUANTILES.

TABLE 4.ONE SAMPLE T-TEST OF RT AND ACCURACY BY GENDER

	Gender	N	Mean	SD	SE
ReactionTime	female	42	2091.649	178.448	27.535
	male	322	2165.405	269.341	15.010
Accuracy%	female	42	0.934	0.064	0.010
	male	322	0.942	0.078	0.004

4.1 Result of RT and Accuracy

Figure 4 shows average reaction times (RT) of correct lane-changes across all conditions with standard error bars. Visual-only tracks indicated the average RT of the first and the second time visual-only tracks (0 and 13). Visual-only tracks mark off the facilitation versus deterioration as a baseline in this experiment. In the present thesis, given that I had clear hypotheses for RT and accuracy, I conducted planned comparisons using paired samples t-tests. For planned comparisons, familywise Type I error rate is generally deemed unnecessary (Keppel & Wickens, 2004). 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 over the visual-only track.

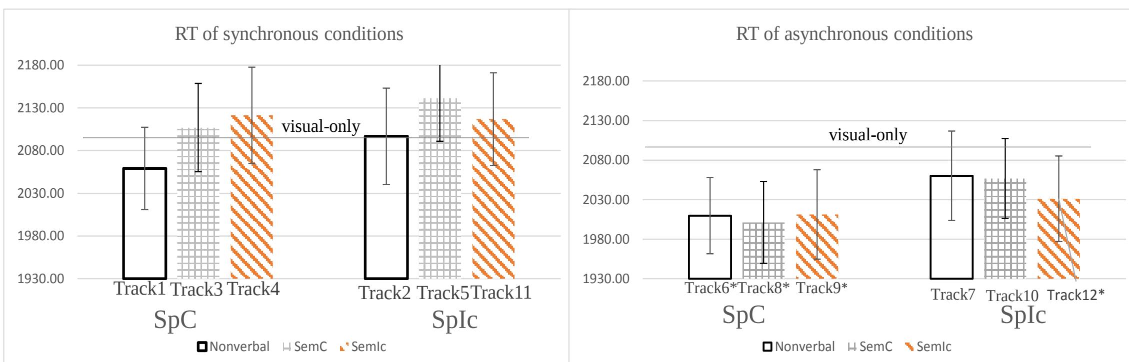


FIGURE 4. BAR PLOT OF RTS IN SYNCHRONOUS VERSUS ASYNCHRONOUS CONDITIONS. NOTES: THE ERROR BARS ARE STANDARD ERRORS. THE ASTERISKS INDICATE A SIGNIFICANT DIFFERENCE FROM VISUAL-ONLY RT.

For tracks with nonverbal cues, Asyn-SpC (Track 6) showed significantly faster RT than visual-only ($t(26) = -2.383$, $p = 0.025$). For tracks with verbal cues, Asyn-SpC-SemC (Track 8), Asyn-SpC-SemIc (Track 9) and Asyn-SpIc-SemIc (Track 12) showed significantly faster RT than visual-only ($t(25) = -2.478$, $p = 0.02$, $t(25) = -2.817$, $p = 0.009$, $t(25) = -2.665$, $p = 0.013$ respectively).

4.2 Result of lane-change accuracy versus secondary task accuracy

Figure .5 is the accuracy of lane-change in twelve conditions versus visual-only baseline. In terms of accuracy, only Syn-SpIc-SemC (Track 5) showed significantly higher accuracy than visual-only ($t(25) = -2.271$, $p = 0.032$). All detailed paired samples t-test results were listed in Appendix.A. I plotted accuracy of the lane change task versus accuracy of secondary auditory memory task (see Figure.6). No correlation was found.

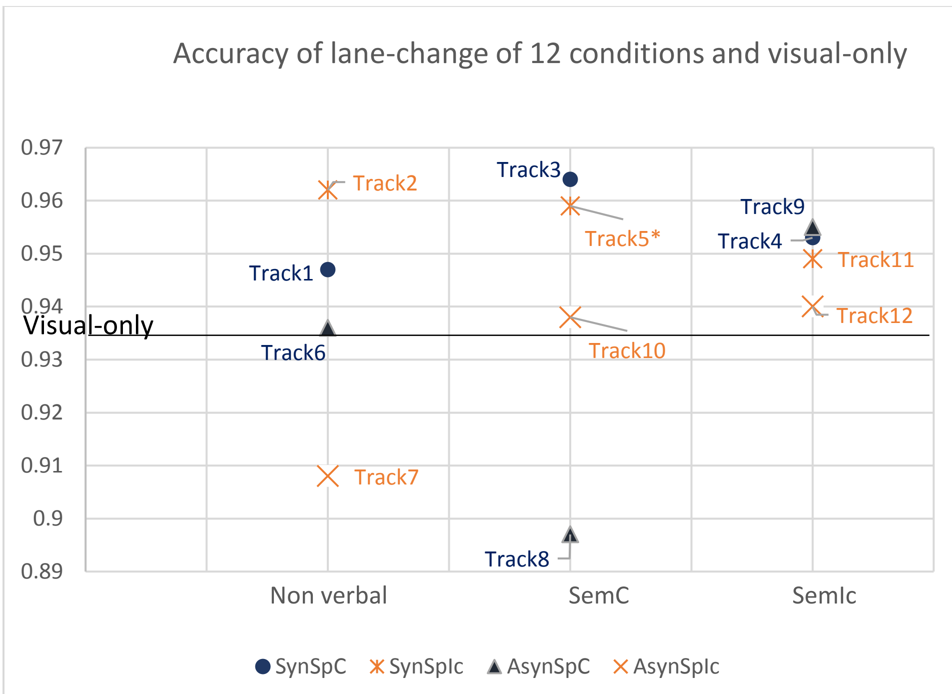


FIGURE 5. ACCURACY OF LANE CHANGE IN TWELVE MULTIMODAL CONDITIONS

For accuracy, there was no clear results or patterns, but synchronous conditions tended to show higher accuracy than asynchronous conditions.

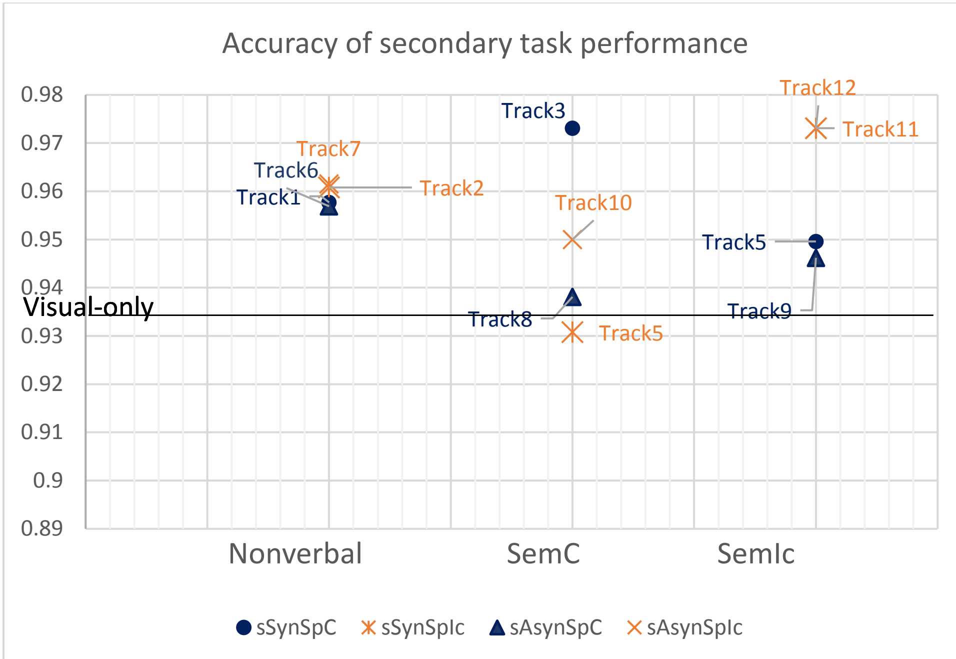


FIGURE 6. ACCURACY OF SECONDARY TASK PERFORMANCE OF TWELVE MULTIMODAL CONDITIONS

4.3 Result of Mdev

Similar to the result of RT and accuracy, I conducted planned comparisons using paired samples t-tests. Only track 8 (AsynSpCSemC) and track 10 (AsyncSpIcSemC) showed signifcantly different lane deviation ($t(25) = 2.095$, $p = .047$). No other significant result was found.

In sum, four out of six asynchronous track showed significant fast RT than visual-only conditions, which suggest priming auditory cue facilitate RT. Only Syn-SpIc-SemC (Track 5) showed significantly higher accuracy than visual-only. But synchronous conditions tended to show higher accuracy than asynchronous conditions.

5. DISCUSSION

5.1. Discussion of RT

The present experiment used the Auditory-Spatial Stroop paradigm (Pieters, 1981) 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. Tables 5, 6, 7 and 8 listed hypotheses and results with a check mark or cross mark to indicate whether the hypotheses were supported or not.

Since visual-only tracks served as the baseline in comparison with all conditions, the subtraction of multimodal tracks over the visual-only tracks were denoted as ΔRT and Δ% in Table 5. The labels "ΔRT" and "Δ%" respectively represent the differences in RT and accuracy between multimodal tracks and visual-only tracks. This simplified version of the twenty four paired-t-test results were used in the discussion section.

TABLE 5. SUBTRACTION OF CONDITIONAL RTS AND ACCURACY OUT OF BASELINE RTS AND ACCURACY

		Nonverbal Cue		**Verbal Cue**			
		SpC	SpIc	SpC		SpIc	
				SemC	SemIc	SemC	SemIc
		Track 1	Track 2	Track 3	Track 4	Track 5	Track 11
Syn	ΔRT	-32.31	5.28	15.51	50.06	29.82	25.46
Syn	Δ%	1.20%	2.70%	2.90%	1.80%	2.50%*	1.40%
		Track 6	Track 7	Track 8	Track 9	Track 10	Track 12
Asyn	ΔRT	-81.64*	-31.17	-90.16*	-80.11*	-34.69	-60.27*
Asyn	Δ%	0.10%	-2.70%	-3.70%	2.00%	0.30%	0.50%

Notes: *$p < 0.05$

5.1.1 Spatial rule (H1)

TABLE 6. SUMMARY RESULT FOR HYPOTHESIS 1 ON SPATIAL RULES

	Nonverbal Cue		**Verbal Cue**			
	SpC	SpIc	SpC		SpIc	
			SemC	SemIc	SemC	SemIc
Syn	Track 1	Track 2	Track 3	Track 4	Track 5	Track 11
	< visual-only	> visual-only	> visual-only	> visual-only	> visual-only	> visual-only
Asyn	Track 6	Track 7	Track 8	Track 9	Track 10	Track 12
	<* visual-only	< visual-only	<* visual-only	<* visual-only	< visual-only	<* visual-only
	H1a	H1b	H1a		H1b	

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 partly supported H1a. It demonstrated that spatially congruent A-V association would enhance visuospatial response speed. As with the spatial rules in multimodal facilitation, it is easier to direct one's attentional focus in different sensory modalities to the same spatial location rather than different location (Spence & Driver, 2004). 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 no matter if the sounds were related to the primary driving task or not

(Spence, 2010). Therefore, the arousal effect might somehow cancel out the spatially incongruent cues' plausible delay effects. Overall, the data tend to support H1c as table 5. shown.

5.1.2 Temporal rule (H2)

TABLE 7. SUMMARY OF RESULT FOR HYPOTHESIS 2 ON TEMPORAL RULES

		Nonverbal Cue		**Verbal Cue**			
		SpC	SpIc	SpC		SpIc	
				SemC	SemIc	SemC	SemIc
		Track 1	Track 2	Track 3	Track 4	Track 5	Track 11
H2b	Syn	< visual-only	> visual-only	> visual-only	> visual-only	> visual-only	> visual-only
		Track 6	Track 7	Track 8	Track 9	Track 10	Track 12
H2a	Asyn	<* visual-only	< visual-only	<* visual-only	<* visual-only	< visual-only	<* visual-only

H2a and H2b were concerned with the temporal rules in crossmodal links. As hypothesized in H2a, the asynchronous multimodal pairs (Track 6, 7, 8, 9, 10, &12) showed shorter RTs than the visual-only baseline, except Asyn-SpIc conditions (Track 7 & 10). Therefore, H2a was mostly supported by the results (The two exceptions, Asyn-SpIc Track 7 & 10, were discussed in H3b). Four out of six in asynchronous pairs supported Posner's preparation function theorem that priming auditory cue benefits RTs. However, Posner only used the non-verbal sound for auditory cues and there was no comparison with verbal cues. The present experiment expended the asynchrony benefit to the verbal cues. The asynchronous (200 msec in this experiment) A-V sped up response time either when there was no location-meaning confliction between A-V modalities or when the auditory cues were only spatially congruent with the visual target and semantically incongruent.

I hypothesized in H2b that RTs in synchronous multimodal pairs would not be longer than those in visual-only conditions. Synchronous incongruent pairs (Track 2, 4, 5, & 11) mostly showed numerically longer RTs than visual-only conditions. The trend seems against H2b. Why did cross-modal synesthesia not happen in this experiment? The Colavita visual dominant effect might be the reason. In the speeded audiovisual asynchrony discrimination tasks, Koppen and Spence (2007) investigated the influence of different SOA (Stimulus Onset Asynchrony). To many synchronous AV pairs, the visual cue was actually perceived 12ms faster than the auditory cue which might lead to a prior-entry effect. In sum, generating auditory cues at the same time with visual cues might not have reached the exactly same timing for cross-modal synesthesia.

5.1.3 Spatial-semanticity conflict in verbal cues (H3)

Table 8. Summary of result for hypothesis 3 on spacity semanticity conflict

	Verbal Cue					
	SpC			SpIc		
	SemC		SemIc	SemC		SemIc
	Track 3		Track 4	Track 5		Track 11
Syn	15.51	<	50.06	29.82	>	25.46
	Track 8		Track 9	Track 10		Track 12
Asyn	-90.16*	<	-80.11*	-34.69	>	-60.27*
		H3a	H3b		H3a	

For the tracks having verbal cues, the spatially and semantically congruent groups had the shortest RTs among verbal pairs (Track 3 has a faster RT than Track 4, 5, &11. Track 8 has a faster RT than Track 9, 10, &12). SpIcSemIc pairs (Tracks 11 & 12). They had better performance than spatially incongruent and semantically congruent pairs (Tracks 5 & 10). This is because the spatial and semantic nature within the verbal cues were still consistent with each other when they were both 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. This trend accords with H3a that spatially and semantically consistent pairs would have better performance than conflicted pairs.

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 (in our brain, where information is more rapidly processed than what information in general), but also the temporal dimension seems to have priority and make interaction results.

5.2 Discussion of accuracy

No clear pattern was found in accuracy of lane-change or accuracy of secondary auditory location memory task. Overall, the redundant gain of the auditory cues towards visual target on accuracy cannot be captured in this experiment. As mentioned, this could be explained by the distinction of the two tasks: visual scanning vs. visual tracking. Identifying the visual indication could be the 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 only maximize the time-sharing benefit when primary task competes secondary task in the same resource (i.e., V-v task especially when the secondary task is visual scanning task.) Also, the time-sharing benefit in visual scanning task (reaction time) would be 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 fast does not guarantee better or smoother control of the vehicle. More research needs to be done to explore to what extent these trade-offs could occur (whether it can be bearable or ultimately harms overall performance).

5.3 Discussion in Mdev

In terms of Mdev, it is the first time to use the intersection area to quantify the lateral control. To my best knowledge, no previous study used such calculation. There was no clear trend in the lane deviation result, except for one comparison. In this exceptional case, AsynSpCSemC (track 8) even showed larger lane deviation than AsyncSpIcSemC (track 10). Note that lane deviation results came from the comparison with the average of the visual-only condition. It is not clear if (1) spatially congruent and semantically congruent auditory cues resulted in much better lane change behavior then the visual-only condition and so, led to bigger deviation or (2) the average of the two visual-only might not be the optimal baseline. If (1) is true, then the similar track, which has asynchronous spatially congruent auditory cue in the non-verbal condition would also have shown larger deviation from the visual-only condition than the spatially incongruent condition. That was not the case in the present experiment. In the future experiment, only one visual-only condition can be used by randomization or full counterbalance of the order.

5.4 Limitations

The auditory preemption theory (Wickens & Liu, 1988) reveals that Va configuration helps processing of the IT (relative to Vv configuration), but would actually hurt processing of the OT. It only works when the two V-v targets are placed in a separated angle. In the present experiment, there was no visual distractor or secondary visual input. In addition, the visual target repeatedly appeared in one place. Low visual workload in the primary driving task might have restricted the multimodal facilitation in lane deviation.

6. CONCLUSION

The current thesis evaluated reaction time and accuracy of the lane change test for different types (verbal vs. nonverbal) of auditory cues manipulated for three dimensions (spatial, semantic, and temporal) in the presence of a visual target. The results showed that the application of the multimodal displays (audio-visual) could improve the lane-change -test performance, but also showed that there were myriads of interactions among variables. Results showed that adding auditory cues could help lane change test more in reaction time than accuracy. The temporal dimension seems to be the most influential in performance. That is, preceding auditory cues improved reaction time. This is in line with Posner's preparation function theorem that the priming audio ahead of the visual target can result in the faster response than the visual-only (temporal rule). Spatially and semantically congruent auditory cues facilitated reaction time. However, RT benefit on ipsilateral cues over contralateral cues for auditory cues (spatial rule) was only supported by the asynchronous pairs. When spatial and semantic 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 (2009) research. However, as the auditory preemption theory (Wickens, Dixon & Seppelt, 2005) suggested, asynchronous A-V cues did not improve accuracy. Only when the spatially incongruent verbal cue appeared simultaneously with the visual target, it hurt accuracy. Moreover, when there is conflict between auditory cues and visual target, having consistency in auditory cues would be more important than having inconsistency within the auditory cue for partial consistency with the visual cue. For example, even though the auditory cue is both spatially and semantically incongruent with visual targets (e.g., “LEFT” verbal cue from the left side with a visual target for the right), if there is an internal consistency between spatial and semantic property within the verbal cues, it was better than the conflict between vision and audition (e.g., “RIGHT” verbal cue from the left side with a visual cue for the right). Also, in-vehicle technology designers would 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. Also, 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 (Jeon, 2016). Then, the model will be able to better explain and predict the effects of non-verbal auditory displays of the multimodal interfaces. Another theoretical point is that part of the results showed sound’s strong arousal effect, which can be better explained by the auditory preemption theory (Wickens, Dixon, & Seppelt, 2005). Certainly, more research is required to disentangle the various influences of auditory cues.

In future studies, it would be interesting to see whether auditory cues can relieve dual-task workload caused by demanding visual scanning secondary task. Given that Posner’s experiment using the 200 msec interval was not 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 reaction timer will be helpful in the maneuver level driving task compared with the operational level (go/no-go) driving task. A similar study using a higher fidelity simulator could also be conducted, which provides a more realistic driving environment. It would help guide in-vehicle technology designers design the system more safely and effectively.

7. REFERENCE

Arend, U., & Wandmacher, J. (1987). On the generality of logical recoding in spatial interference tasks. *Acta Psychol (Amst), 65*(3), 193.

Baldwin, C. L. (2012). *Auditory cognition and human performance: Research and applications*: CRC Press.

Barrow, J. H., & Baldwin, C. L. (2009). *Verbal-spatial cue conflict: implications for the design of collision-avoidance warning systems.* Paper presented at the Proceedings of the International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Broadbent, D. E. (1958). *Perception and communication*: Oxford University Press.

Chan, A. H., & Or, C. K. (2012). A comparison of semantic and spatial stimulus–response compatibility effects for human–machine interface design. *European Journal of Industrial Engineering, 6*(5), 629-643.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*(2), 409-412.

Deutsch, J. A., & Deutsch, D. (1963). Attention: some theoretical considerations. *Psychol Rev, 70*(1), 80.

Engström, J., Arfwidsson, J., Amditis, A., Andreone, L., Bengler, K., Cacciabue, P. C., Janssen, W. (2004). *Meeting the challenges of future automotive HMI design: an overview of the AIDE integrated project.* Paper presented at the European Congress on Intelligent Transportation Systems and Services, 4th, 2004, Budapest, Hungary.

Engström, J., & Markkula, G. (2007). *Effects of visual and cognitive distraction on lane change test performance.* Paper presented at the Proceedings of the 4th international driving symposium on human factors in driver assessment, training, and vehicle design.

Ho, C., & Spence, C. (2012). *The multisensory driver: Implications for ergonomic car interface design*: Ashgate Publishing, Ltd.

ISO, I. (2010). 26022: 2010 Road vehicles–Ergonomic aspects of transport information and control systems–Simulated lane change test to assess in-vehicle secondary task demand. Norm. *International Organization for Standardization, Geneva, Switzerland, 24*.

Keppel, G., & Wickens, T. (2004). Design and Analysis, A Researcher's Handbook. 2004: Pearson/Prentice Hall, Upper Saddle River, New Jersey.

Koppen, C., & Spence, C. (2007). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain research, 1186*, 224-232.

Liu, Y.-C., & Jhuang, J.-W. (2012). Effects of in-vehicle warning information displays with or without spatial compatibility on driving behaviors and response performance. *Applied Ergonomics, 43*(4), 679-686.

Mattes, S. (2003). *The Lane Change Task as a Tool for driver Distraction Evaluation. IHRA-ITS Workshop on Driving Simulator Scenarios, October 2003-Dearborn, Michigan. wwwnrd. nhtsa. dot. gov*. Retrieved from

Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do? *Human behavior and traffic safety* (pp. 485-524): Springer.

Jeon, M. (2016). *How is nonverbal auditory information processed? Revisiting existing models and proposing a preliminary model.* Paper presented at the Human Factors and Ergonomics Society Annual Meeting, Washington, DC.

Olsheski, J. D. (2014). The role of synesthetic correspondence in intersensory binding: investigating an unrecognized confound in multimodal perception research.

Petzoldt, T., Brüggemann, S., & Krems, J. F. (2014). Learning effects in the lane change task (LCT)–Realistic secondary tasks and transfer of learning. *Applied Ergonomics, 45*(3), 639-646.

Pieters, J. M. (1981). Ear asymmetry in an auditory spatial Stroop task as a function of handedness. *Cortex, 17*(3), 369-379.

Posner, M. I., Klein, R., Summers, J., & Buggie, S. (1973). On the selection of signals. *Memory & cognition, 1*(1), 2-12.

Proctor, R. W., Tan, H. Z., Vu, K.-P. L., Gray, R., & Spence, C. (2005). *Implications of compatibility and cuing effects for multimodal interfaces.* Paper presented at the Proceedings of the HCI International 2005.

Proctor, R. W., & Van Zandt, T. (2008). *Human factors in simple and complex systems*: CRC press.

Proctor, R. W., & Vu, K.-P. L. (2006). *Stimulus-response compatibility principles: Data, theory, and application*: CRC Press.

Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Consciousness and cognition, 17*(3), 989-1015.

Sinnett, S., Soto-Faraco, S., & Spence, C. (2008). The co-occurrence of multisensory competition and facilitation. *Acta Psychol (Amst), 128*(1), 153-161.

Spence, C. (2010). Crossmodal spatial attention. *Ann N Y Acad Sci, 1191*(1), 182-200.

Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention*: Oxford University Press.

Spence, C., & Soto-Faraco, S. (2010). Auditory perception: interactions with vision. *The Oxford handbook of auditory science: Hearing, 3*, 271-296.

Sterkenburg, J. (2015). Impacts of distraction on driving: An analysis of physical, cognitive, and emotional distraction.

Tattegrain, H., Bruyas, M.-P., & Karmann, N. (2009). *Comparison Between Adaptive and Basic Model Metrics in Lane Change Test to Assess In-Vehicle Secondary Task Demand.* Paper presented at the PROCEEDINGS OF THE 21ST (ESV) INTERNATIONAL TECHNICAL CONFERENCE ON THE ENHANCED SAFETY OF VEHICLES, JUNE

2009, STUTTGART, GERMANY. http://www-nrd.nhtsa.dot.gov/Pdf/ESV/esv21/09-0252.pdf (retrieved 08/30/16)

Trbovich, P. L. (2006). *Effects of phonological, visual and spatial information processing on a simulated driving task.* Carleton University Ottawa.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental brain research, 181*(1), 173-181.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & cognition*, *13*(4), 304-319.

Wickens, C. D., Dixon, S. R., & Seppelt, B. (2005). *Auditory preemption versus multiple resources: who wins in interruption management?* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Wickens, C. D. (2008). Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 449-455.

Wickens, C., Prinet, J., Hutchins, S., Sarter, N., & Sebok, A. (2011). *Auditory-Visual Redundancy in Vehicle Control Interruptions Two Meta-analyses.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering Psychology and Human Performance*: Pearson Education, Limited.

Workload. (2016, July 31). In Wikipedia, The Free Encyclopedia. Retrieved 02:31, August 12, 2016, from https://en.wikipedia.org/w/index.php?title=Workload&oldid=732396591

Young, K. L., Lenné, M. G., & Williamson, A. R. (2011). Sensitivity of the lane change test as a measure of in-vehicle system demand. *Applied Ergonomics*, *42*(4), 611-618.

Zylstra, B., Tsimhoni, O., Green, P., & Mayer, K. (2003). Driving performance for dialing, radio tuning, and destination entry while driving straight roads. *Ann Arbor, MI: The University of Michigan Transportation Research Institute*.

Appendix. A

TABLE 9.PAIRED SAMPLES T-TEST ON RT AND ACCURACY

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Track1 - V26	-32.31	114.81	22.52	-78.69	14.06	-1.435	25	.164
Pair 2	Track6 - V26	-81.64	174.70	34.26	-152.21	-11.08	-2.383	25	.025*
Pair 3	Track2 - V26	5.28	183.04	35.90	-68.65	79.21	.147	25	.884
Pair 4	Track7 - V26	-31.16	136.14	26.70	-86.15	23.83	-1.167	25	.254
Pair 5	Track3 - V26	15.51	193.16	37.88	-62.51	93.53	.409	25	.686
Pair 6	Track8 - V26	-90.16	185.49	36.38	-165.08	-15.24	-2.478	25	.020*
Pair 7	Track4 - V26	50.06	154.51	30.30	-12.35	112.47	1.652	25	.111
Pair 8	Track9 - V26	-80.11	145.02	28.44	-138.69	-21.54	-2.817	25	.009*
Pair 9	Track5 - V26	29.82	153.23	30.05	-32.07	91.71	.992	25	.331
Pair 10	Track10 - V26	-34.69	113.67	22.29	-80.60	11.22	-1.556	25	.132
Pair 11	Track11 - V26	25.46	145.53	28.54	-33.33	84.24	.892	25	.381
Pair 12	Track12 - V26	-60.27	115.29	22.61	-106.83	-13.70	-2.665	25	.013*
Pair 13	Track1 - V26acc	.01	.07	.01	-.02	.04	.826	25	.416
Pair 14	Track6 - V26acc	.00	.10	.02	-.04	.04	.055	25	.957
Pair 15	Track2 - V26acc	.03	.08	.02	.00	.06	1.774	25	.088
Pair 16	Track7 - V26acc	-.03	.11	.02	-.07	.02	-1.287	25	.210
Pair 17	Track3 - V26acc	.03	.08	.02	.00	.06	1.789	25	.086
Pair 18	Track8 - V26acc	-.04	.10	.02	-.08	.00	-1.931	25	.065
Pair 19	Track4 - V26acc	.02	.07	.01	-.01	.05	1.302	25	.205
Pair 20	Track9 - V26acc	.02	.08	.02	-.01	.05	1.301	25	.205
Pair 21	Track5 - V26acc	.02	.06	.01	.00	.05	2.271	25	.032
Pair 22	Track10 - V26acc	.00	.07	.01	-.02	.03	.243	25	.810
Pair 23	Track11 - V26acc	.01	.07	.01	-.02	.04	.971	25	.341
Pair 24	Track12 - V26acc	.01	.08	.02	-.03	.04	.345	25	.733
Pair 25	visual0 - V26	75.04	119.69	23.47	26.70	123.39	3.197	25	.004
Pair 26	visual0 - V26acc	-.04	.05	.01	-.06	-.02	-4.099	25	.000
Pair 27	Track1 – Track2	-37.598	211.582	41.495	-123.058	47.862	-.906	25	.374
Pair 28	Track6 - Track7	-50.483	147.109	28.850	-109.902	8.935	-1.750	25	.092
Pair 29	Track3 – Track5	-14.31	130.48	25.59	-67.01	38.397	-0.559	25	0.581
Pair 30	Track8- Track10	-55.47	228.39	44.79	-147.7	36.773	-1.238	25	0.227
Pair 31	Track4 - Track11	24.6	142.26	27.9	-32.85	82.059	0.882	25	0.386

Pair 32	Track9 - Track12	-19.85	140.27	27.51	-76.5	36.805	-0.722	25	0.477

Notes. The results from 26 participants in four orders. (The number of participants in each groups are 9, 8, 4, and 5). The first two orders are reverse with each other and the third and fourth orders are reverse with each other. The alpha level was 0.05. V26 means the average of two visual-only tracks. Visual0 means the first-time visual-only tracks.