In-vehicle air gesture design: impacts of display modality and control orientation

In-vehicle air gesture design: impacts of display modality and control orientation

In-vehicle Air Gesture Design: Impacts of Display Modality and Control Orientation

Jason Sterkenburg[1], Steven Landry[1], S. Maryam FakhrHosseini[2], and Myounghoon Jeon[3]

[1]Department of Cognitive and Learning Sciences, Michigan Technological University, Houghton, MI, USA

[2]AgeLab, Massachusetts Institute of Technology, Cambridge, USA

[3]Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

*Corresponding author: Myounghoon Jeon Tel.: +1-540-231-3510, E-mail: myounghoonjeon@vt.edu

Keywords: air gesture system, direct manipulation, driver workload, dual task paradigm

In-vehicle Air Gesture Design: Impacts of Display Modality and Control Orientation

Abstract

The number of visual distraction-caused crashes highlights a need for non-visual displays in the in-vehicle information system (IVIS). Audio-supported air gesture controls can tackle this problem. Twenty-four young drivers participated in our experiment using a driving simulator with six different gesture prototypes–3 modality types (visual-only, visual/auditory, and auditory-only) X 2 control orientation types (horizontal and vertical). Various data were obtained, including lane departures, eye glance behavior, secondary task performance, and driver workload. Results showed that the auditory-only displays showed a significantly lower lane departures and perceived workload. A tradeoff between eyes-on-road time and secondary task completion time for the auditory-only display was also observed, which means the safest, but

slowest among the prototypes. Vertical controls (direct manipulation) showed significantly lower workload than horizontal controls (mouse metaphor), but did not differ in performance measures. Experimental results are discussed in the context of multiple resource theory and design guidelines for future implementation.

1. Introduction

The prevalence of touchscreens in vehicles has increased in recent years (Hassel, 2016). Touchscreen use in vehicles introduces a conflict for visual attention between driving and in-vehicle information system (IVIS) use. This conflict has been shown to increase crash risk (Dingus et al., 2006; Horrey & Wickens, 2007; Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006; Olson, Hanowski, Hickman & Bocanegra, 2009; Wierwille & Tijerina, 1998) and has been a subject of concern among driving researchers for many years (Green, 2000; Ranney, Mazzae, Garrott, & Goodman, 2000; Burnett, Summerskill, & Porter, 2004) which has sparked efforts to develop new IVISs that reduce the demands for drivers' visual attention (e.g., Sodnik et al., 2008; Reiner, 2012; May, Gable, & Walker, 2014; Shakeri, Williamson, & Brewster, 2017).

Recent technological advances have made it possible to cheaply and effectively measure hand positions of drivers using infrared sensors (e.g., LEAP Motion) or computer vision technologies (e.g., Microsoft Kinect). Researchers have explored these technologies as an effective means to develop in-vehicle control systems that are easier to use and reduce the crash risk associated with using traditional IVISs (May, Gable & Walker, 2014; Gable, Raja, Samuels & Walker, 2015). Fundamentally, the operation of air gesture controls described here is similar to the current touchscreen model. Inputs are still based on the WIMP (windows, icons, menus, pointer) style of interaction, i.e., users select menu items laid out in a hierarchy via control of a cursor. This is opposed to a symbolic system controlled via performance of dynamic gestures such as taps, swipes, or a type of sign language.

To develop an air gesture control system that is less visually demanding than touchscreens, auditory displays can be used to convey information about cursor position (e.g.,

Sterkenburg, Landry & Jeon, 2019). Well-designed air gesture controls supported by auditory displays could supplement or even replace the visual information needed to use an IVIS, allowing drivers to focus visual attention on the road while operating in-vehicle controls eyes-free.

The goal of this study is to understand the effects of auditory cues and display-control orientation compatibility on driving performance, secondary task performance, eye glance behavior, and perceived workload. To the best of our knowledge, this is the first experiment employing an auditory-only in-vehicle gesture display. We are particularly interested in comparing the *auditory-only* display with the visual-only display and the visual/auditory display. Also, this is the first experiment testing the compatibility between drivers' arm movement orientation and display control in the air gesture navigation task. We are curious to see whether vertical movement, which represents a direct manipulation (Shneiderman, 1997) or horizontal movement, which presents a lower physical workload, would be a more effective orientation for arm movement. We expect that this experiment will practically contribute to the design of sonically enhanced in-vehicle gesture systems. The results of the present study will also theoretically contribute to advancing multimodal interaction and its tradeoff and to applying the basic HCI principle (e.g., direct manipulation) to the actual interaction design in the vehicle setting.

2. Related Work

2.1 Impacts of In-vehicle Controls on Driver Distraction

Driver distraction is defined as competition leading to diversion of attention away from driving to secondary tasks that results in degraded driving (Young & Regan, 2007). Under this definition of driver distraction, IVIS use is a driver distraction. The task left to driving

researchers and IVIS designers is to mitigate the crash risk associated with IVIS and reduce the probability of a crash to the lowest possible level. One of the theories that can explain this dynamic is multiple resource theory (MRT). Based on MRT, people have a limited set of resources to process the information around themselves (Basil, 1994). MRT has four orthogonal resource dimensions. Stage includes perception, cognition, and response selection. Modality includes visual and auditory. Code includes verbal and spatial. Visual type includes focal vision and peripheral vision. In practice, MRT can predict task performance by accounting for variability in task interference and concurrently performed tasks (Wickens, 2002). For example, while a driver is driving, they can also listen to music because each task is using different resources (visual and auditory). The utility of MRT in this pursuit is that MRT can predict that when multi-tasking has to be conducted using different resources (i.e., modalities in this study), it leads to better time-sharing between the tasks (Wickens, 2002).

MRT provides a solid basis for explaining driver distraction and the cognitive perspectives of driving task which involves visual perception, manual manipulation, and spatial coding of environment. In this regard, considerable focus on driver distraction has been related to the use of cell phones in vehicles. Meta-analysis of 28 experiments on texting and driving showed that texting increases off-road eye glances, reaction times to changes in the environment, number of collisions, and vehicle headway, and reduces lane control and speed (Caird, Johnston, Willness, Asbridge, & Steel, 2014). Another meta-analysis of 23 studies on the effects of talking on a cell phone while driving showed that cell phones primarily degrade driving by increasing reaction times, rather than reducing lane control (Horrey & Wickens, 2006).

Infotainment systems also require visual demands. Tijerina and colleagues examined distractions associated with route guidance systems (Tijerina, Parmer, & Goodman, 1998). They found that destination entry in a route guidance system took substantially longer to complete than cell phone dialing or tuning a radio. They also found that visual-manual inputs took longer, increased the number of off-road glances and number of lane departures compared to a voice-controlled system. Naturalistic observations of drivers using different route guidance methods, i.e., paper maps, route guidance without voice guidance, and route guidance with voice guidance, revealed that both conventional maps and route guidance without voice guidance resulted in increased visual demands and driving degradation (Dingus et al., 1995; Srinivasan & Jovanis, 1997). Route guidance systems with voice guidance were associated with the best performance.

When touchscreen technology was introduced to vehicle head units, researchers began to focus on touchscreen keyboards and their impacts relative to voice command technology (Tsimhoni, Smith, & Green, 2004). Results showed that touchscreen keyboards took longer to use than voice inputs, and also degraded lane keeping more than voice input controls. Touchscreens also include more complicated WIMP-inspired ("windows, icons, menus, pointer") interfaces, which introduce layers of menu depth, and require precise movements, and searching for and selecting small targets that are grouped closely together, as in toolbar or ribbon menus (Balakrishnan, 2004; McGuffin & Balakrishnan, 2005). As a consequence, touchscreen use may require more visual demand compared to other methods of in-vehicle control use. Additionally, both driving and in-vehicle controls require biomechanical resources, which, in combination with visual demands (e.g., text entry into route guidance systems), have

been shown to degrade driving performance (Hurwitz & Wheatly, 2002; Tijerna, Palmer, & Goodman, 1998).

In all of the above examples, drivers got distracted by secondary and tertiary tasks that often are highly demanding and use similar resources or exhaust the available ones. Both driving (primary) task and other tasks require manual spatial responses and focal vision. Consequently, drivers might fail to divide their attention and allocate their resources properly, which leads to distraction and slower response time to the driving tasks. Voice can be an alternative input interface. However, it is still not widely accepted in the automotive domain because of technical difficulties, continuous control, and visibility of command feedback (e.g., Goulati & Szostak, 2011; Pfleging, Schneegass, & Schmidt, 2012; Pickering, Burnham, & Richardson, 2007). In-vehicle gesture interactions may outperform voice input interfaces while requiring less precise movement (so, lower demanding of manual resources than touchscreen) and less focal vision. See section 2.4 Air Gesture Controls in Vehicles for more details.

2.2 Eye Glances and Driving

The driving literature clearly points to conflict for visual attention as one of the major causes of distraction-related crashes. Peng et al. (2013) showed in a naturalistic study that drivers' ability to maintain good lane control degrades proportionately with the eyes-off-road-time. Donmez et al. (2010) showed that drivers who had non-visual feedback completed tasks on their infotainment systems while driving without looking away from the road as frequently compared to using the system with only visual feedback. In addition, according to NDS (naturalistic driving study) data taken from real-world drivers by Klauer et al. (2006), short glances away from the road pose little or no risk to driving safety compared to a baseline condition in which drivers drove with no imposed distraction. But long glances away from the

road–2 seconds or more –increase near-crash/crash risk by at least two times normal driving (Klauer et al., 2006).

2.3 In-vehicle Auditory Displays

Auditory displays have been frequently used in devices designed for visually-impaired individuals (Gaver, 1989; Edwards, 1989; Mynatt & Edwards, 1992). Auditory displays have also been shown to decrease subjective workload and improve performance for sighted users completing computer-based drag and drop tasks as well (Brewster, 1998a; Brewster, 1998b). We can consider that drivers are temporarily visually impaired for non-driving tasks because their vision is heavily taxed on the road. Indeed, the meta-analytic studies have demonstrated that auditory displays or multimodal displays that provide visual and auditory information outperform visual-only displays in vehicles (Wickens & Seppelt, 2002; Liu, 2001). For example, Jeon and colleagues (Jeon, Gable, Davison, Nees, Wilson, & Walker, 2015) showed that the use of auditory displays with visual cues significantly enhanced driving performance as well as the secondary menu navigation performance and perceived workload in a driving context. Research has shown that auditory displays led to relatively better performance using an in-vehicle gesture system (Shakeri, Williamson, & Brewster 2017; Sterkenburg et al., 2019). As Sterkenburg et al. showed, speech appears to offer an easy path to differentiation and identification, which makes it a potential design element to include within an in-vehicle information system. The above observations once again can be explained by MRT by addressing the resources involved in these tasks; driving demands visual resources (not auditory resources), so auditory displays compete less with driving than do visual displays.

2.4 Air Gesture Controls in Vehicles

If drivers are required to move their hands over the surface of the screen to search and navigate through the menu, it will require significant hand-on-touchscreen time. It requires not only driver hands-off from the wheel (i.e., manual spatial responses), but also causes driver visual distraction because touchscreen control demands a driver's focal vision. Currently, the J287 SAE standard provides guidelines that detail where to place controls in vehicles so that most people can reach them and use them (Society for Automotive Engineers, 1988; 2007). However, more recent research has shown that these reach envelope standards may allow for reachable controls but they are not necessarily easily reachable, and some of the limits are at medium difficulty levels on average for drivers (Yu et al., 2017; Liu et al., 2017). Of course, auditory/tactile displays on touchscreens could still be a viable solution, especially if positioned in a more easily reachable position. However, because either auditory displays or tactile displays are feedback, it still requires drivers' visual attention and if not, they have to touch the item first and move on to the next item. Meanwhile, gesture sensors can record movement data within a wide range of space, allowing for less physically demanding reaching movements for drivers. Also, drivers can hover over the item before they select the menu item. Air gesture controls also require driver hands-off from the wheel. However, research has shown that adding in-vehicle gesture interfaces with well-designed auditory displays improved visual distraction while leading to equivalent driving performance compared to touchscreen (Sterkenburg, Landry, & Jeon, 2019) or significantly improved lane deviation and steering wheel angle depending on the menu design (Tabbarah, 2022).

There are many questions surrounding the application of air gestures in vehicles. As a result, there have been many different types of research done on this topic. Research has focused on the engineering of the software and hardware required for air gestures to work (Akyol, Canzler, Bengler, & Hahn, 2000; Ohn-bar, Tran, & Trivedi, 2012), some has focused on pointing

gestures (Cairnie, Ricketts, Mckenna & Mcallister, 2000) or static symbolic gestures (Aykol et al., 2000), and others on motion-path gestures (Rahman, Saboune, Saddik & Ave, 2011). Most of the studies have either not developed a gesture control system (Alpern & Minardo, 2003) in favor of Wizard-of-Oz methodologies or they have not conducted any evaluation of system usability or its impact on driving (Akyol et al., 2000; Cairnie et al., 2000; Rahman et al., 2011). In this study, we both developed and evaluated a working prototype air gesture control system.

Despite the demand for eyes-free in-vehicle controls, there is little work for which researchers have developed air-gesture controls and evaluated the system's usability and impact on driving performance. One exception comes from May, Gable, and Walker (2014) who performed an experiment in which participants drove in a simulator while completing simple menu navigation tasks using both air gesture controls and touchscreens. They found that driving performance was comparable between the two systems, but air gesture control actually resulted in more short glances away from the road and participants reported a higher overall workload. Despite mixed results, eye glance behavior was still within NHTSA guidelines (National Highway Traffic Safety Administration, 2012).

In a different study, 20 participants volunteered for an end-user gesture and voice elicitation experiment (Bilius, & Vatavu, 2020). Drivers' preferences for gesture and voice input collected using a 5-point Likert scale. Results showed that the majority of drivers considered gesture input useful. Moreover, drivers were asked about the car functions they would like to control using gesture. Window control, aerator mouth, and head-up display are among the top three ranked functions. In addition to all the in-vehicle tasks and functions, Jiang, Xia, Liu, and Bai, (2020) proposed to replace most used touch operations with gesture controls for mobile devices in a driving environment. They validated their design with three case studies.

Regarding the type of gesture controls for navigating interfaces, Wu, Gable, May, Choi, and Walker (2016) examined four gesture interaction techniques. They found that drivers experience higher workload with air gestures than surface gestures. On the other hand, compared to traditional air-conditioning control, Jahani, Alyamani, Kavakli, Dey, and Billinghurst, (2017) showed that mid-air gestures reduce driving errors by up to 50%.

Another exception comes from Shakeri, Williamson, and Brewster (2017) who evaluated the impacts of different display modalities on lane deviations, eye glance behavior, and secondary task performance. They found that auditory displays outperformed tactile displays for secondary task performance, but performed *worse* than the visual display condition. However, the auditory displays led to drastically reduced eyes-off-road-time. Regarding driving performance, there were no differences in observed lane deviations.

One potential benefit of gesture controls is the ability to utilize three-dimensional space, which allows for more efficient use of space. However, the utility of three-dimensional space is not easily realized in vehicles because three-dimensional menus could be too demanding physically and cognitively to be operated while driving. The objective of air gestures and, likewise, touch gestures, is to obtain the safe and effective use of in-vehicle information systems. Bach, Jaeger, Skov, and Thomassen (2008) showed in their research that use of non-visual touch gesture interfaces did not result in improvements relative to traditional touchscreen interfaces with regard to driving safety or performance. Instead, their touch gesture interface demonstrated reduced visual demand, as intended, but at the cost of degraded performance using the interface, i.e., drivers took longer to complete tasks using the gesture interface but they did not need to look away from the road as frequently.

The potential advantages of an air-based gesture control system over a touch-based system remains an open question. The recent study (Sterkenburg et al., 2019) showed that auditory-supported air gestures allowed drivers to look at the road more, showed equivalent driver workload and driving performance, but slightly decreased secondary task performance compared to touchscreens.

3. The Current Study and Hypotheses

3.1 Air Gesture System Design

The purpose of this study was to learn about the impacts of gesture control orientation and the combinations of visual and auditory displays. To this end, we designed a 2x2 grid menu, with only four square targets, 5x5 inches across (Sterkenburg et al., 2019; Figure 1). The visual display shows a grid, with the menu item name in each box. The visual display also shows the cursor position, represented by a small colored box, and also highlights each menu item box in white whenever the cursor is in it. When a selection is made the visual display changes the highlight color to indicate to the driver they have made a selection. These design decisions were made to visually convey as much information as possible to the driver so they can gather information at a glance (highlighted box) or in detail (cursor position) and so they have confidence that the system is responding to them (cursor and selection highlight).

The selection gesture, i.e., the gesture that drivers make to select a menu item was an open hand. This choice was made to mitigate, as much as possible, the number of false positives from the LEAP Motion sensor. The system occasionally miscounts the number of visible fingers. The best way to reduce the frequency of miscounts was to require the system to see five fingers to make a selection. That way, the driver can keep their hand closed and the system will be very unlikely to count five fingers. The drawback of this selection gesture is that the center of the

palm, which determines the cursor position, moves as a consequence of the hand-opening movement. On balance, this gesture still seemed to be more beneficial than harmful considering the limitations of the spatial resolution limitations of the LEAP Motion sensor (approximately 1 cm error), and its tendency to miscount fingers.

There were two types of metaphor regarding the movement plane–mouse control and direct manipulation. In the metaphor of a mouse movement with a computer, the participants' movement plan was a horizontal plane. In other words, if the participant moves along the Y axis, the cursor moves along the Z axis (i.e., moving forward to move up the cursor in the display). We expected that this orientation would be less physically demanding than movements on the vertical plane. In the metaphor of direct manipulation, the participants' movement plan followed a vertical plane. That is, if the participant moves along the Z axis, the cursor moves along the Z axis (i.e., moving upwards to move up the cursor in the display). The movement on the X axis was same in the two conditions.

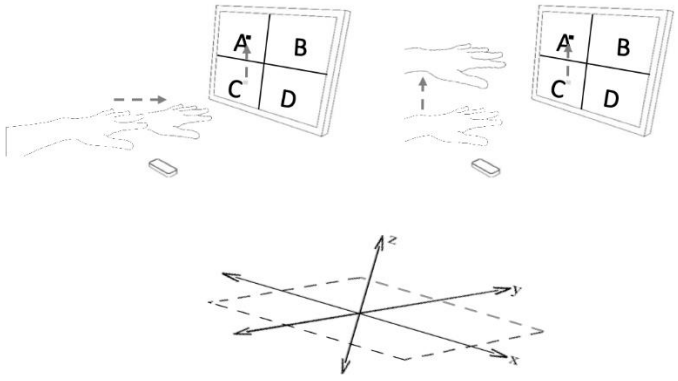

Figure 1. Gesture system and control orientation. In the horizontal condition (left), a participant moved forward (Y axis) to move up on the menu (Z axis). In the vertical condition (right), the participant moved upward (Z axis) to move up on the menu (Z axis)

3.2 Hypotheses

H1: Auditory-only condition will lead to slower and less accurate target selections than visual-only and visual-auditory conditions (May et al., 2014; Sterkenburg et al., 2019).

H2: The frequency of off-road glances will be lower in the conditions with auditory-only and visual-auditory conditions than the visual-only condition. The introduction of auditory displays will reduce visual demand of the secondary task (Shakeri et al., 2017; Sterkenburg et al., 2019).

H3: The average following distance to the lead vehicle will be highest for the auditory-only condition, compared to the visual-auditory and visual-only conditions. The variance in following distance will be higher for the visually demanding conditions. Variance in following distance will be highest for visual-only condition and lowest for the auditory-only condition (Strayer & Drew, 2004). See Section 4.5 Procedure for the detailed task description.

H4: Vertical menu will lead to higher physical demand, compared to the horizontal menu orientation. However, the vertical menu will lead to a higher percentage of correct selections when compared to the horizontal menu because of direct manipulation (Grossman & Balakrishnan, 2004)

4. Methods

4.1 Participants

A total of 24 undergraduate psychology students (21 males, 3 females) were recruited for this within-subject design experiment (Table 1). All participants were given course credit as compensation for their participation. Only one person reported having experience using a LEAP Motion before. All participants were required to have at least 1.5 years of driving experience to control for the confounding variables due to inexperienced driving. After each participant

reviewed and signed the consent form approved by Institutional Review Board (IRB), a moderator started the study with the training session.

Table 1. Participant demographics

	Age (yrs)	Experience (yrs)	Miles/yr
Mean	19.67	3.5	6540
SD	0.96	1.15	7033

4.2 Experimental Design

We used a within-subjects study design. There was a total of six conditions (Table 2). There were three levels of visual/audio display: visual, audio, and visual/audio. There were also two levels of control orientation: vertical and horizontal. With a fully orthogonal design, there were six conditions.

Table 2. Experimental conditions

	Auditory	Visual/Auditory	Visual
Vertical	VA	VVA	VV
Horizontal	HA	HVA	HV

4.3 Apparatus and Stimuli

A LEAP Motion was used as our hand-position tracking sensor. To develop our target selection task, we used Pure Data–an open source graphical programming language. As the participant moved their hand above the sensor, a cursor matched the position of the person's hand along the X, Y, or Z axes and made corresponding movements on the screen (Figure 1). All cursor movements were mapped one-to-one to hand movements. For the prototypes that

have auditory feedback, the recorded male voice (e.g., saying “B”) was used when hovering over the menu item. A selection action was followed by a confirmatory sound, which contained two “raindrop” tones, the first low note followed immediately by a second higher frequency note. This was intended to provide an indication of selection.

A National Advanced Driving Simulator MiniSim medium-fidelity driving simulator (Figure 2) was used for all driving scenarios. The simulator consisted of three Panasonic TH-42PH2014 42" plasma displays, each with a 1280x800 pixel resolution, which allowed 130-degree field of view in front of the seated participant. The center monitor was 28 inches from the center of the steering wheel and the left and right monitors were 37 inches from the center of the steering wheel. The MiniSim also included a real steering wheel, adjustable car seat, gear-shift, and gas and brake pedals, as well as a Toshiba Ltd. WXGA TFT LCD monitor with a 1280x800 resolution to display the speedometer, etc. The driving scenario consisted of a single closed circuit through a residential area with many left and right curves. With the exception of the lead vehicle, there were no other cars in the scenario. Participants were asked to drive between 30-40 mph over the duration of the experiment. The simulator automatically recorded the following distance to the lead vehicle, lane position and vehicle speed.

As seen in the Figure 2, the gesture control system was positioned to the right of the driver sitting in the driving simulator. The center of the monitor was positioned 16 inches from the right edge of the steering wheel. The angle of the monitor was not strictly controlled, but was angled slightly to improve visibility to drivers. The sensor position was also fixed in position 12 inches from the right edge of the steering wheel.

Figure 2. Driving simulator setup, visual display monitor with webcam, and LEAP Motion

4.4 Dependent Measures

Speed: average speed in miles per hour and standard error of speed were recorded.

Lane departures – percentage of drive duration where at least one tire has departed from the lane boundaries. This is measured by the distance of the center of the driver's vehicle from the center of the correct lane. Whenever the vehicle strayed more than 4.0 meters from the center of the lane, the vehicle was considered outside of the correct lane.

Eye glance behavior: number of glances of three different durations: short (<1 second), medium (1-2 seconds), and long (>2 seconds). Eye glance behaviors were recorded by a webcam placed on top of the visual display monitor (Figure 2). The eye glances were later coded by a researcher and placed into three categories based on the estimated length of the glance duration above. We chose these categories because NHTSA guidelines state that at least 85% of off-road eye glances should be less than two seconds (National Highway and Traffic Safety Administration, 2012).

Secondary task performance: movement time in milliseconds marks the duration between the cue prompting participants to start a movement and a correct selection. Selection accuracy is defined by the percentage of selections that are made correctly.

Driver workload: The NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988) is a subjective assessment tool that provides a standardized measure of workload. It consists of six primary scales that participants use to rate their perceived workload. The questions in the NASA-TLX assessment are as follows:

Mental Demand: "How mentally demanding was the task?" This scale measures the level of mental effort, complexity, and cognitive requirements experienced by the participant while performing the task.

Physical Demand: "How physically demanding was the task?" This scale assesses the amount of physical effort and activity required by the participant to complete the task.

Temporal Demand: "How hurried or rushed was the pace of the task?" This scale evaluates the extent to which the participant feels rushed or pressed for time during the task.

Performance: "How successful do you think you were in accomplishing the task?" This scale reflects the participant's perception of their performance success or accomplishment during the task.

Effort: "How hard did you have to work to accomplish your level of performance?" This scale measures the level of effort exerted by the participant to achieve their perceived performance level.

Frustration: "How insecure, discouraged, irritated, stressed, and annoyed were you?" This scale assesses the level of negative feelings or frustration experienced by the participant during the task.

The scales are rated using a continuous numerical scale. The rating ranges and definitions are as follows:

Mental Demand, Physical Demand, Temporal Demand, Effort, and Frustration: These scales are rated on a scale from 0 to 100. The lower end of the scale (0) represents "very low" or "none," indicating minimal demand or frustration. The higher end of the scale (100) signifies "very high" or "extremely," indicating maximum demand or frustration.

Performance: This scale is rated on a scale from 0 to 100, where 0 represents "completely unsuccessful" or "total failure," and 100 represents "completely successful" or "perfect performance."

Participants assigned ratings on each scale based on their subjective experience and perception of the task workload. The final NASA-TLX score was derived by summing up the ratings across all six scales. Higher total scores indicate higher perceived workload experienced by the participant.

4.5 Procedure

After the consent form procedure, participants were trained to use the gesture control systems for five minutes. This time was spent training on each of the different conditions, approximately one minute for each condition. This ensured that none of the conditions was new to a participant during the test session. This training was done to mitigate as much as possible the learning effects associated with using a totally novel air gesture control system. Participants then practiced driving in the simulator for several minutes to become acclimated and even practiced using the menu system and driving simultaneously. The participants were given no instructions about how they should balance the demands of the primary and secondary tasks. As a primary task, they were instructed to follow the lead vehicle, maintaining the same distance until the end of the driving scenario.

This scenario was chosen to gather insights into the drivers' attention and distraction levels, particularly in terms of how their speed and distance vary while interacting with different interfaces and tasks.

The order in which participants used the prototypes was counterbalanced in a Latin Square design such that each condition appears in each position in the order. This design washed out order effects associated with the learning curve of using an air gesture control system. A total of 32 selection tasks, evenly divided between target options, were completed for each prototype system, taking approximately five minutes to complete. Speech cues instruct participants which target to select (e.g., “Select Navigation”). The order of the auditory cues was randomly determined by the Pure Data patch.

After completing all of the selection tasks, notes were taken about participants’ first impressions. Next, participants were asked several questions about their workload (Hart & Staveland, 1988) including: mental demand, physical demand, temporal demand, performance, effort, and frustration from the NASA-TLX workload assessment. After rating each subscale, they chose the scale title that represents the more important contributor to workload for the task between the two subscale pairs. This process was repeated for all six prototypes.

4.6 Statistics

Repeated-measures ANOVAs (3x2 within-subjects design) were conducted to measure the effects of two factors on driving performance, secondary task performance, and workload: Display, Orientation. Two-tailed, paired-samples t-tests were conducted when factors with three or more levels showed a significant difference. However, if a significant three-way interaction existed, all pairs were compared. A Holm-Bonferroni correction was applied to decrease the number of Type-1 errors. This correction lowers the critical p-value from 0.05 to 0.017 for the

Display Factor, but remains at 0.05 for the Orientation factor. Partial eta squared was also reported as a measure of effect size. For the secondary task measure's time and accuracy, one participant's data were removed from analysis because data were missing due to experimenter error.

5. Results

5.1 Lane Departures

Lane departures are defined by the percentage of time during which at least a part of the vehicle is outside the correct lane. Repeated measures ANOVA results showed no significant effect of Orientation on lane departures, $F(1,23) = 0.058$, $p = 0.812$, $\eta_p^2 = 0.002$. The Display factor did show a significant effect, $F(2,46) = 4.437$, $p = 0.017$, $\eta_p^2 = 0.162$. There were no statistically significant interactions between factors, for Orientation and Display $F(2,46) = 0.696$, $p = 0.504$, $\eta_p^2 = 0.029$. Pairwise comparisons showed fewer lane departures for Auditory displays than Visual displays, $t(23) = 3.168$, $p = 0.008$, but there were no significant differences between Auditory displays and Visual/Auditory displays, $t(23) = 2.220$, $p = 0.063$, and Visual/Auditory displays and Visual displays, $t(23) = -1.828$, $p = 0.074$ (Figure 3).

Another measure of lane control is standard deviation of lane position, which is a measure of swerving on the road while driving. ANOVA results showed no statistical significance for the main effect of Orientation on standard deviation of lane position, $F(1,23) = 2.411$, $p = 0.134$, $\eta_p^2 = 0.095$. The Display factor showed a significant main effect on standard deviation of lane position, $F(2,46) = 10.83$, $p < 0.001$, $\eta_p^2 = 0.320$. There were no statistically significant interactions between Orientation and Display, $F(2,46) = 1.093$, $p = 0.344$, $\eta_p^2 = 0.045$. Pairwise comparisons showed significantly lower standard deviation of lane deviations for the Auditory displays compared to the Visual displays, $t(23) = 5.120$, $p < 0.001$, and Visual/Auditory

displays, $t(23) = 2.967$, $p = 0.009$. The Visual display was not statistically different from Visual/Auditory displays, $t(23) = -2.203$, $p = 0.033$ (Figure 4).

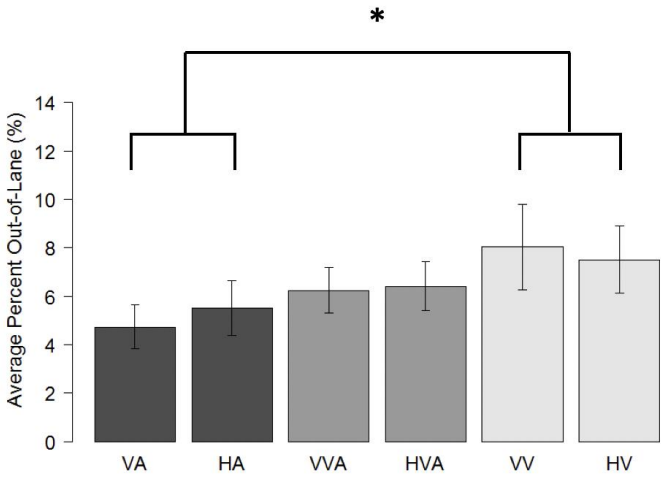

Figure 3. Average percentage of time spent out-of-lane (VA: Vertical Auditory; HA: Horizontal Auditory; VVA: Vertical Visual Auditory; HVA: Horizontal Visual Auditory, VV: Vertical Visual; HV: Horizontal Visual)

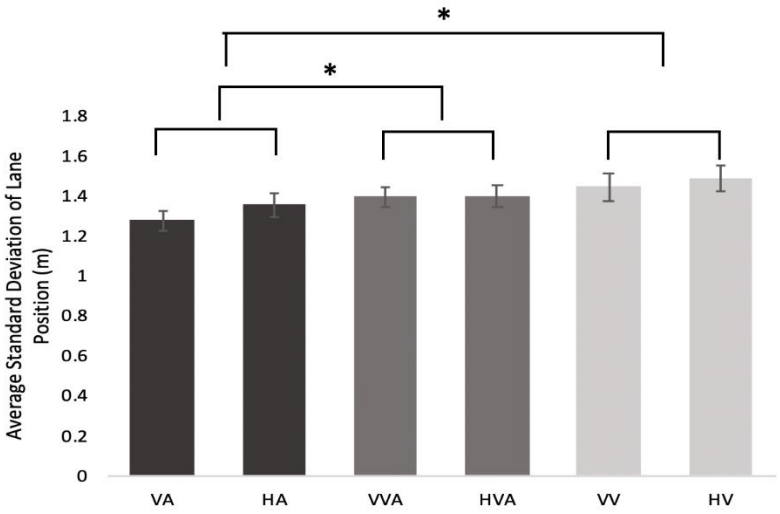

Figure 4. Average percentage of time spent out-of-lane (VA: Vertical Auditory; HA: Horizontal Auditory; VVA: Vertical Visual Auditory; HVA: Horizontal Visual Auditory, VV: Vertical Visual; HV: Horizontal Visual)

The presence of a visual display was shown to degrade driving performance by increasing the amount time spent out of the lane and by also leading to increased standard deviation of lane position. Conditions with auditory-only displays had significantly lower standard deviations in lane position and a lower percentage of drive outside of the correct lane, as shown by the paired t-tests. For standard deviation of lane position, the auditory-only condition led to improvements even over the visual/auditory displays. Meanwhile, the orientation of the control had no impact on lane control.

5.2 Following Distance

Repeated Measures ANOVA results showed no significant main effect for the Orientation factor on mean following distance, $F(1,23) = 0.005$, $p = 0.947$, $\eta_p^2 < 0.001$. The Display factor did have a significant effect on mean following distance, $F(2,46) = 4.702$, $p = 0.014$, $\eta_p^2 = 0.178$. There were no statistically significant interactions between Orientation and Display $F(2,46) = 1.474$, $p = 0.24$, $\eta_p^2 = 0.061$. Paired comparisons showed significantly greater mean following distance for the Visual displays compared to the Auditory displays, $t(23) = 3.505$, $p = 0.003$. But there were no significant differences between Visual and Visual/Auditory displays, $t(23) = -1.692$, $p = 0.195$, or Auditory and Auditory/Visual displays, $t(23) = 0.962$, $p = 0.341$ (Table 3).

ANOVA results showed no significant effect of Orientation on standard deviation of following distance, $F(1,23) = 0.480$, $p = 0.496$, $\eta_p^2 = 0.004$. There was also no statistically significant effect of the Display factor on standard deviation of following distance, $F(2,46) = 1.479$, $p = 0.239$, $\eta_p^2 = 0.062$. There was also no statistically significant interaction between Orientation and Display, $F(2,46) = 2.272$, $p = 0.115$, $\eta_p^2 = 0.092$.

Table 3. Means and standard deviations for following distance from lead vehicle

	HA	HV	HVA	VA	VV	VVA
mean (m)	130.77	140.21	129.28	121.12	130.54	128.67
sd (m)	44.56	52.23	45.01	46.64	45.17	45.49

Overall, the mean following distance was reduced by the addition of an auditory display, but no other factors impacted mean following distance to a statistically significant level (Table 4). The standard deviation of following distance showed no main effects from any of the factors, but tend to be increased by removing the auditory display when using a horizontal control orientation (i.e., HVA vs. HV).

5.3 Eye Glances

5.3.1 Short glances (<1 seconds)

ANOVA results showed there was no significant main effect of Orientation on the number of short eye glances, $F(1,23) = 3.198$, $p = 0.087$, $\eta_p^2 = 0.122$. Display did show a significant main effect on the number of short off-road eye glance, $F(1,23) = 39.58$, $p < 0.001$, $\eta_p^2 = 0.632$. There were no significant statistical interactions between the Orientation and Display factors, $F(1,23) = 2.382$, $p = 0.136$, $\eta_p^2 = 0.094$.

Pairwise comparisons showed fewer off road eye glances for the Auditory conditions compared to Visual displays, $t(23) = 19.031$, $p < 0.001$, and Visual/Auditory displays, $t(23) = 7.315$, $p < 0.001$. The Visual displays led to more off-road eye glances compared to the Visual/Auditory displays, $t(23) = -10.783$, $p < 0.001$ (Table 4 and Figure 5).

Overall, these results showed that the presence of both visual and auditory displays impacted the number of short off-road eye glances. The addition of auditory displays clearly decreased the number of off-road eye glances while the addition of visual displays led to an

increase in the number of off-road eye glances. These effects had very large effect sizes and can be seen in Figure 5.

5.3.2 Medium glances (1-2 seconds)

ANOVA results showed there was no effect of Orientation on the number of medium off-road eye glances, $F(1,23) = 2.35$, $p = 0.139$, $\eta_p^2 = 0.093$. Displays showed a main effect on the number of medium off-road eye glances, $F(1,23) = 20.04$, $p < 0.001$, $\eta_p^2 = 0.466$. There were no statistically significant interactions between Orientation and Display $F(1,23) = 2.353$, $p = 0.139$, $\eta_p^2 = 0.093$.

Paired samples t-tests showed significantly fewer medium off-road eye glances for the Auditory conditions compared to the Visual conditions, $t(23) = 6.705$, $p < 0.001$, and Visual/Auditory displays, $t(23) = 4.868$, $p < 0.001$. The Visual condition resulted in more medium off-road eye glances compared to Visual/Auditory conditions, $t(23) = -5.452$, $p < 0.001$ (Figure 5).

5.3.3 Long glances (>2 seconds)

ANOVA results showed no significant effect of the control Orientation factor on the number of long eye glances, $F(1,23) = 0.063$, $p = 0.802$, $\eta_p^2 = 0.003$. The Display factor showed a significant effect on long off-road eye glances, $F(1,23) = 5.697$, $p = 0.026$, $\eta_p^2 = 0.199$. There were no statistically significant interactions between Orientation and Display, $F(1,23) = 0.057$, $p = 0.814$, $\eta_p^2 = 0.002$. Pairwise comparisons showed fewer long off-road eye glances for Auditory displays compared to Visual displays, $t(23) = 2.808$, $p = 0.022$. But there were no significant differences between Visual displays and Visual/Auditory displays, $t(23) = -2.185$, $p = 0.055$, or Auditory displays and Visual/Auditory displays, $t(23) = 2.280$, $p = 0.055$ (Table 4).

Table 4. Means and standard deviations of off-road glance counts across conditions

		HA	HV	HVA	VA	VV	VVA
Short	mean	1.750	25.625	13.917	0.833	25.043	10.000
	sd	2.707	7.966	10.413	1.239	10.052	10.100
Medium	mean	0.000	3.875	1.583	0.000	3.174	0.958
	sd	0.000	3.069	1.863	0.000	4.075	1.546
Long	mean	0.000	0.458	0.167	0.000	0.565	0.208
	sd	0.000	1.141	0.482	0.000	1.376	0.658

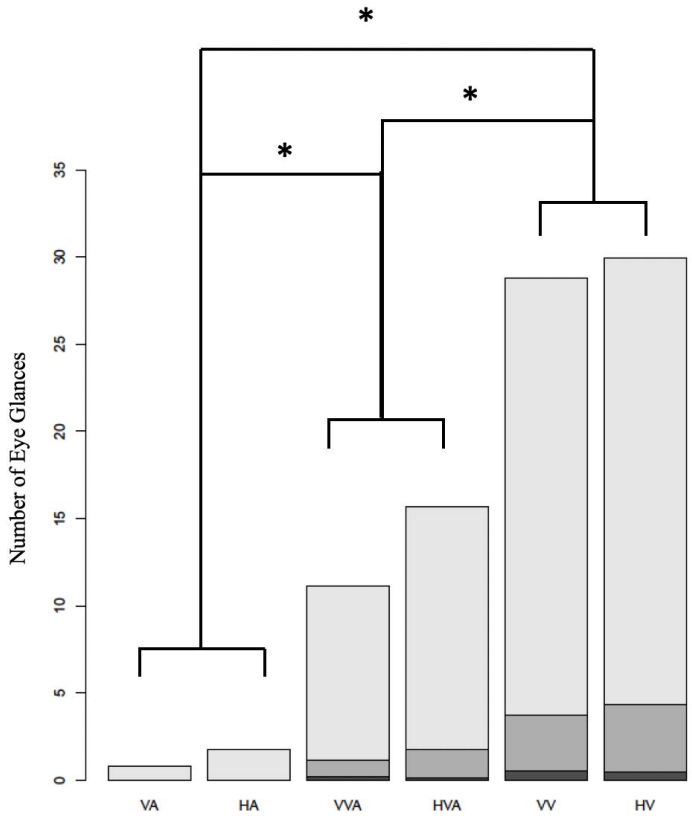

Figure 5. Eye glance frequency for short (light grey), medium (grey), and long glances (dark grey)

5.4 Menu Selection Time

ANOVA results showed no significant effect of Orientation on selection times, $F(1,22) = 0.778$, $p = 0.387$, $\eta_p^2 = 0.034$. The Display condition had a significant impact on selection times,

$F(2,44) = 23.93$, $p < 0.001$, $\eta_p^2 = 0.521$. There was no statistically significant interaction between Orientation and Display, $F(2,44) = 0.097$, $p = 0.908$, $\eta_p^2 = 0.004$. Pairwise t-tests showed significant differences between all combinations of displays: Visual/Auditory displays were slower than Visual, $t(22) = 2.550$, $p = 0.014$, but faster than Auditory displays, $t(22) = -5.389$, $p < 0.001$. Auditory displays were slower than Visual displays, $t(22) = -7.333$, $p < 0.001$ (Table 5).

Table 5. Means and standard deviations of selection times for the secondary task

	VA	**HA**	**VVA**	**HVA**	**VV**	**HV**
mean (ms)	3213	3307	2848	2846	2672	2736
standard error (ms)	267	335	244	253	230	242

5.5 Menu Selection Accuracy

ANOVA results showed no effect of Orientation on task completion accuracy, $F(1,22) = 0.875$, $p = 0.36$, $\eta_p^2 = 0.038$. The Display factor also had no significant effect on task accuracy, $F(2,44) = 0.3$, $p = 0.742$, $\eta_p^2 = 0.013$. There were no statistically significant interactions between Orientation and Display, $F(2,44) = 0.571$, $p = 0.569$, $\eta_p^2 = 0.025$. All conditions resulted in mean accuracy rates between 88-92% (Table 6).

Table 6. Means and standard deviations for secondary task accuracy

	HA	**HV**	**HVA**	**VA**	**VV**	**VVA**
Mean	88.7%	91.3%	91.9%	92.5%	92.9%	91.5%
Standard Error	6.5%	5.8%	5.6%	5.4%	5.2%	5.7%

5.6 Perceived Workload

The NASA Task Load Index (NASA-TLX) relies on self-reporting. The scale ranges from zero to 100, where zero indicates very low workload, and 100 signifies very high workload.

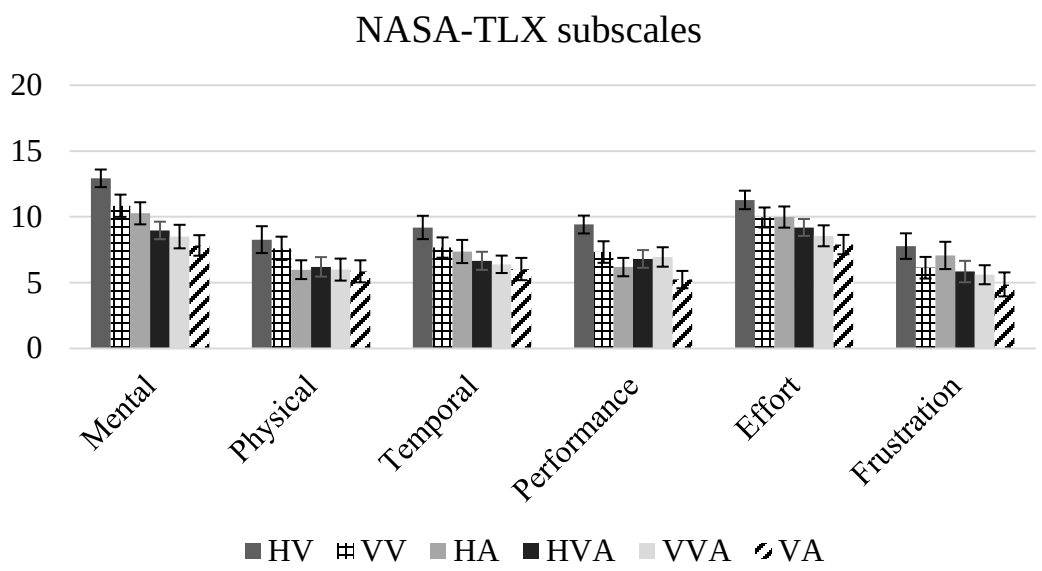

Figure 6. NASA-TLX workload subscales

5.6.1 Mental workload

Participants reported their mental workload on a scale of 0 to 100 by responding to the following question from the NASA-TLX index: "how mentally demanding was the task?" ANOVA results showed a significant main effect for Orientation on mental demand, $F(1,23) = 10.76$, $p = 0.003$, $\eta_p^2 = 0.319$ (Figure 6). The Display factor also showed a significant main effect on mental demand, $F(2,46) = 18.66$, $p < 0.001$, $\eta_p^2 = 0.632$. There was no significant interaction between Display and Orientation factors, $F(2,46) = 2.299$, $p = 0.112$, $\eta_p^2 = 0.091$. Pairwise t-tests showed the Visual conditions led to significantly higher perceived mental workload compared to Visual/Auditory, $t(23) = 5.565$, $p < 0.001$, and Auditory conditions, $t(23) = -5.574$, $p < 0.001$. The Auditory conditions led to similar perceived mental demand compared to the Visual/Auditory displays, $t(23) = 0.587$, $p = 0.560$.

5.6.2 Physical workload

Participants reported their physical demand on a scale of 0 to 100 by responding to the following question from the NASA-TLX index: "how physically demanding was the task?" ANOVA results showed no effect of Orientation on physical workload, $F(1,23) = 0.43$, $p = 0.519$, $\eta_p^2 = 0.018$ (Figure 6). The Display factor had a main effect on physical workload, $F(2,46) = 4.944$, $p = 0.011$, $\eta_p^2 = 0.177$. There were no significant statistical interactions between Orientation and Display, $F(2,46) = 0.193$, $p = 0.825$, $\eta_p^2 = 0.008$. Pairwise t-tests showed the Visual conditions led to significantly higher perceived physical workload compared to Visual/Auditory, $t(23) = 3.580$, $p = 0.002$, and Auditory conditions, $t(23) = -2.904$, $p = 0.011$. The Auditory conditions led to similar perceived physical demand compared to the Visual/Auditory displays, $t(23) = -0.314$, $p = 0.755$.

5.6.3 Temporal workload

ANOVA results showed no significant effect of Orientation on temporal workload, $F(1,23) = 3.933$, $p = 0.059$, $\eta_p^2 = 0.146$ (Figure 6). The Display factor showed a main effect on temporal workload, $F(2,46) = 7.993$, $p = 0.001$, $\eta_p^2 = 0.258$. There were no statistically significant interactions between Orientation and Display, $F(2,46) = 1.17$, $p = 0.319$, $\eta_p^2 = 0.048$. Pairwise t-tests showed the Visual conditions led to significantly higher perceived temporal workload compared to Visual/Auditory, $t(23) = 4.225$, $p < 0.001$, and Auditory conditions, $t(23) = -3.386$, $p = 0.003$. The Auditory conditions led to similar perceived temporal demand compared to the Visual/Auditory displays, $t(23) = 0.726$, $p = 0.353$.

5.6.4 Performance

Participants rated their performance on a scale of 0 to 100 by responding to the following question from the NASA-TLX index: "how successful were you in accomplishing what you were asked to do?" ANOVA results showed no significant effect of Orientation on performance, $F(1,23)$

= 2.878, p = 0.103, η_p^2 = 0.111 (Figure 6). The Display factor showed a significant effect on performance, $F(2,46)$ = 12.67, p < 0.001, η_p^2 = 0.355. There were no statistically significant interactions between Orientation and Display, $F(2,46)$ = 2.268, p = 0.115, η_p^2 = 0.090. Paired t-tests showed significantly better perceived performance for the Auditory conditions compared to the Visual conditions, $t(23)$ = -4.983, p = 0.001, and Visual/Auditory conditions, $t(23)$ = -2.304, p = 0.026. The Visual/Auditory conditions were lower than the Visual conditions, $t(23)$ = 2.655, p = 0.022.

5.6.5 Effort

Participants rated their effort on a scale of 0 to 100 by responding to the following question from the NASA-TLX index: "how hard did you have to work to accomplish your level of performance?" ANOVA results showed significant main effects for Orientation, $F(1,23)$ = 6.876, p = 0.015, η_p^2 = 0.230 (Figure 6). The Display factor also showed a significant main effect on perceived effort, $F(2,46)$ = 7.708, p = 0.001, η_p^2 = 0.251. There were no statistically significant interactions between Orientation and Display, $F(2,46)$ = 0.746, p = 0.48, η_p^2 = 0.031. Paired samples t-tests showed significantly higher effort for the Visual conditions compared to Visual/Auditory, $t(23)$ = 3.11, p = 0.010, and Auditory conditions, $t(23)$ = -2.941, p = 0.010. Auditory conditions and Visual/Auditory conditions were statistically equivalent, $t(23)$ = 0.120, p = 0.905.

5.6.6 Frustration

Participants rated their frustration on a scale of 0 to 100 by responding to the following question from the NASA-TLX index: "how insecure, discouraged, irritated stressed, and annoyed were you?" ANOVA results showed no significant effect for Orientation, $F(1,23)$ = 4.044, p = 0.056, η_p^2 = 0.150 (Figure 6). The Display factor showed a significant effect on frustration, $F(2,46)$

= 2.375, p = 0.104, η_p^2 = 0.094. There were no statistically significant interactions between Orientation and Display, $F(2,46)$ = 1.745, p = 0.186, η_p^2 = 0.071. But the pairwise t-tests showed no significant differences between Visual/Auditory and Visual prototypes, $t(23)$ = 2.370, p = 0.066, the Visual/Auditory and Auditory, $t(23)$ = 0.414, p = 0.681, or the Visual and Auditory system, $t(23)$ = -1.638, p = 0.216.

5.6.7 Overall workload

The overall workload scale is an overall score that is calculated based on the raw subscale scores and a weight variable assigned to each subscale based on paired ratings in which participants answered which among each pair of subscales contributed more to their workload. ANOVA results showed a significant effect of Orientation on overall workload, $F(1,23)$ = 9.884, p = 0.005, η_p^2 = 0.301 (Figure 7). The Display factor also showed a significant effect on overall workload, $F(2,46)$ = 15.05, p < 0.001, η_p^2 = 0.396. There were no statistically significant interactions between Orientation and Display, $F(2,46)$ = 2.175, p = 0.125, η_p^2 = 0.086. Pairwise t-tests showed the Visual conditions led to significantly higher perceived overall workload compared to Visual/Auditory, $t(23)$ = 5.045, p < 0.001, and Auditory conditions, $t(23)$ = -5.037, p < 0.001. The Auditory conditions led to similar perceived overall workload compared to the Visual/Auditory displays, $t(23)$ = -0.076, p = 0.939.

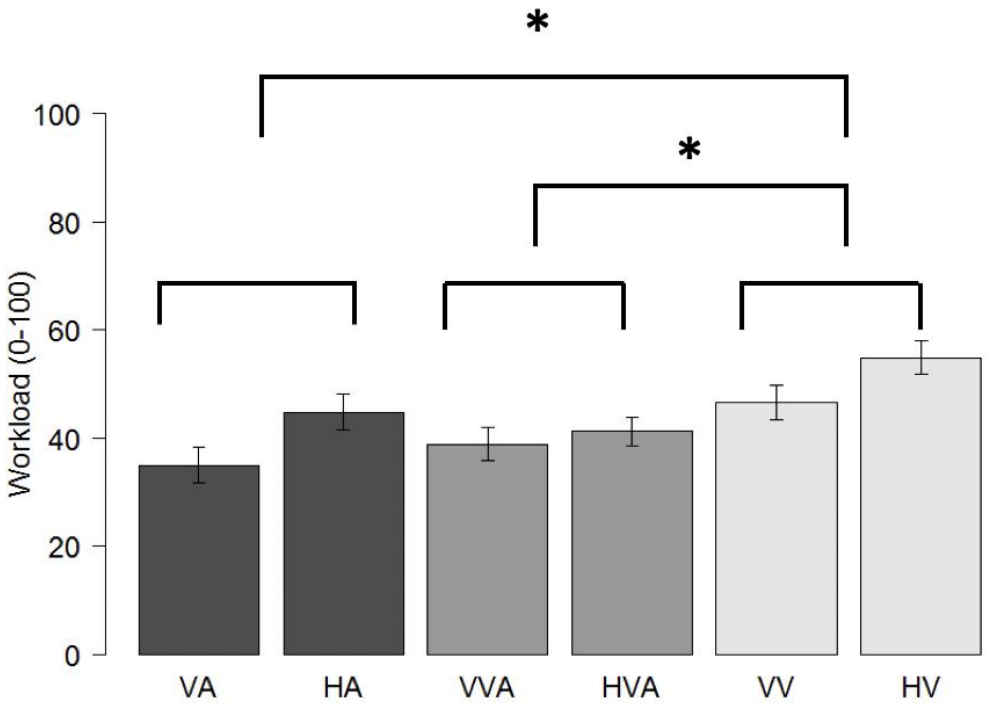

Figure 7. overall workload scores for each condition

6. Discussion

This experiment aimed to investigate the influences of two main factors of in-vehicle air gesture control design–display modality and control orientation–on driving performance, eye glance behavior, secondary task performance, and workload. The results showed that display modality influenced driving performance (lane departures, standard deviation of lane position, and following distance), but control orientation had no effects on performance. In the case of display modality, there was a consistent pattern demonstrating that auditory-only displays led to better driving performance compared to visual-only conditions–fewer lane departures, and lower standard deviation of lane position. This is consistent with the expectation of Multiple Resource Theory which suggests that the addition of an auditory display should allow drivers to process

auditory information to complete the secondary task rather than compete with the driving task for visual processing resources.

The improvement in lane departures associated with auditory-only displays is inconsistent with previous literature which has shown that auditory-supported in-vehicle air gesture systems lead to similar lane deviations as air gesture controls with visual displays (Shakeri et al., 2017). One explanation for the inconsistency is that the driving task in the study from Shakeri et al. (2017) required the driver to drive in a straight line, whereas the driving scenario from our experiment required adapting to changes in speed from a lead vehicle and also adapting to curves on the road. The added difficulty might make driving performance metrics more sensitive to the differences in visual attention demands of secondary tasks. In other words, the visual demand to drive in a straight line is lower than the visual demand to adapt to a lead vehicle and a curvy road. This could explain why driving performance was actually improved with auditory-only gesture controls compared to visual-only gesture controls in our experiment.

Regarding eye glance behavior, the display modality factor had significant impacts on the frequency of short, medium, and long off-road eye glances. Conditions with auditory displays resulted in fewer off-road eye glances and conditions with visual displays were associated with increased off-road eye glances. Again, the control orientation factor had little or no impact on the number of off-road eye glances. The impact of auditory and visual displays is consistent with expectation from MRT, which suggests that drivers should be able to look at the road more when the secondary task can be accomplished without focal visual attention, as is the case for prototypes with an auditory display. This result is also consistent with results from Shakeri et al. (2017), which showed that visual-only displays with air gesture controls lead to greater eyes-off-road time compared to auditory-supported air gesture controls.

Secondary task performance showed that conditions with auditory displays led to slower target selections compared to conditions with visual displays, but the display modality had no impact on secondary task accuracy. Again, the control orientation had no significant impact on either secondary task performance measures. This result is consistent with findings from May et al. (2014). The secondary task completion times showed the same pattern, slower completion times using auditory-supported air gesture controls. The selection accuracies were also equivalent between auditory-supported air gesture controls and visual-only air gesture controls in May et al. (2014), which is the same result observed in the present study.

Workload results showed that visual-only displays led to greater mental demand, physical demand, temporal demand, effort, and overall workload compared to the visual/auditory displays and auditory displays. The performance subscale showed greatest performance for the auditory-only condition. The vertical orientation was associated with reduced overall workload, mental demand, and effort.

Hypothesis 1 stated that auditory-only displays would lead to slower and less accurate target selections. These predictions were logical extensions from the observations in Sterkenburg et al. (2019) that conditions with auditory displays led to slower selection times and lower accuracy. In the present study, results showed slower selection times for auditory-only conditions compared to visual-auditory or visual-only conditions. However, the results also showed menu selection accuracy rates were not lower for auditory-only conditions compared to conditions with visual displays. The slower selection times can be explained by the low bandwidth of auditory information in guiding search tasks and the relatively slow uptake of non-visual information in guiding target selections (Elliott, Helsen, & Chua, 2001). The comparable rate of correct target selections suggests that, at least in the case where there are only a small number of large targets

(index of difficulty < 2 bits), non-visual information is sufficient to make accurate selections. In the case of this experiment, participants were able to hear the auditory display speak the name of the target currently being selected. The auditory display design allows participants to get the same information, whether through the visual or auditory modality, i.e., in which target is the cursor right now. In fact, the only additional information provided by the prototypes with visual displays is the more fine-grain position of the cursor within the menu item, which was not necessary for the task.

Hypothesis 2 stated that visual-only displays would lead to more off-road eye glances compared to visual-auditory displays, which would lead to more off-road eye glances than the auditory-only display. This result was expected for all durations of eye glance. Results from this experiment supported this hypothesis.

Hypothesis 3 stated that auditory-only displays would lead to larger following distances from the lead vehicle. This was supported by literature (Strayer & Drew, 2004) which demonstrated that drivers will compensate by allowing larger following distances behind lead vehicles to compensate when completing secondary tasks. The assumption was that the auditory-only prototype would result in the highest workload and would therefore lead to the greatest compensation in the driving task. The second part of the hypothesis was that drivers would have the greatest variance in following distance when using the visual-only displays, followed by visual-auditory, and auditory-only displays. This hypothesis was based on research that showed visual distractions lead to crash risk because of increased reaction times to changes in the driving environment (Klauer, et al., 2006; Horrey & Wickens, 2006). Because the two experiments in Sterkenburg et al. (2019) showed that the prototypes with visual displays led to greater numbers of off-road glances, if the same holds true for the present study, then participants would be more

likely to miss braking events from the lead vehicle, which would result in delayed reaction times and more variable following distance. Regarding the following distance, the data of the present study showed that drivers actually had the smallest mean following distance for the auditory-only display condition, the exact opposite of the expected result. This result begs two possible explanations: 1) the workload felt by the participants when using the auditory-only display was lower than it was expected to be, and 2) it is possible that because drivers were able to keep their eyes on the road while using the auditory display, they felt more confident in their ability to react to braking events from the lead vehicle and therefore, more comfortable following the lead vehicle at a closer mean distance. With regard to the standard deviation of following distance, the data showed no significant main effects for the display modality factor, meaning that the addition of a visual display did neither lead to increased standard deviations in following distance, nor was there any statistically significant difference between the auditory-only and visual-auditory conditions.

Hypothesis 4 stated that drivers would feel greater physical workload when using prototypes with the vertical control orientation and that the prototypes with vertical control

orientations would lead to higher overall accuracy, as was shown in Grossman and Balakrishnan (2004). The results showed that both parts of hypothesis 4 were unsupported. The control orientation had no statistical impact on physical workload. The prediction that the vertical control orientation would be physically more difficult was based on the assumption that participants would raise their arms at the shoulder in order to keep their hand on a parallel plane with the sensor and free from visual obstruction from their arm, sleeve, or wrist. However, this assumption was not supported by actual participants' behavior, because they raised their arms while keeping their elbows low, leading to relatively lower physical demand even with the

vertical orientation. The hypothesis that the vertical orientation would lead to higher secondary task accuracy was also unfounded, as all of the conditions led to statistically equivalent accuracy rates. The hypothesis that the vertical orientation would lead to better accuracy was based on observations from Grossman and Balakrishnan (2008) that showed selection accuracies for movements on the x-axis (left and right) and movements on the z-axis (up and down) (vertical) were higher than movements along the y-axis (forward and backward) (horizontal). This hypothesis was also consistent with Sterkenburg et al. (2019) that some participants struggled to reach menu items on the bottom left because selecting those targets required them to reach slightly behind themselves in a pocket of space that was especially difficult for some participants, whereas the vertical orientation would not require participants to move their hands backward from that same position. The results of this experiment showed that selection accuracies were not influenced by control orientation. This result could be explained by the large target sizes. The large target sizes might make the task easy and so, participants performed similarly well in making accurate selections. The results from the first experiment in Sterkenburg et al. (2019) showed lower accuracy rates for the 4x4 grid and showed only lowest accuracies for menu items in the very lower left corner of the grid. The average selection accuracies for the lower left quadrant (all four menu items in the lower left corner) were highly variable. It is possible that if the menu items were smaller, then accuracy differences between vertical and horizontal control orientations would have manifested themselves. However, smaller targets would lead to lower accuracies and be less viable for use in an in-vehicle gesture control system. The hypothesis regarding the physical demand was also unmet by our data. While it is still possible that there may be differences in the physical demands of movements along the y-axis and z-axis, the task requirements were relatively low (large targets) leading to fast selection

times, less movement, and more recovery time for the participants. Interestingly, the horizontal control led to significantly higher overall workload, mental demand, and effort than the vertical control. We can cautiously posit that the mouse movement metaphor (horizontal) was more cognitively demanding than the direct manipulation (vertical).

7. Conclusion and Future Work

Overall, the results suggest that the addition of auditory displays led to improved driving performance, less eyes-off-road time, and lower workload across every subscale. Meanwhile visual displays led to improved secondary task performance but degraded driving performance, and led to more off-road eye glances. Orientation had very little impact on behavioral metrics, but did impact perceptions of workload, with participants strongly preferring the vertical orientations. During a short post-experiment interview, participants were asked about each of the experimental factors. The most frequently cited reason for preferring the vertical orientation was its intuitive mapping to the corresponding visual display. The biggest drawback of auditory displays is that they appear to require a longer time to make secondary task selections. This should come as no surprise, given the greater bandwidth afforded by the visual modality in comparison to auditory modality. As supported by results from this experiment, the addition of auditory displays presents a tradeoff between driving safety and efficient secondary task performance.

In future work, we would implement more realistic menu systems (e.g., combinations of one dimensional list menu + grid menu) in one user interface to obtain more external validity. More diverse auditory displays will be tested in addition to speech (e.g., earcons, auditory icons, spearcons) (Walker, Lindsay, Nance, Nakano, Palladino, Dingler, & Jeon, 2013). Speech is the most obvious auditory menu display, but other non-speech auditory displays can overcome the

length of speech cues. In the present study, we used a web cam to observe the participants' eye glances. However, we plan to conduct the next study with the eye-tracking glasses. It will refine the eye glance movement analysis depending on areas of interest with higher accuracy. As mentioned, the LEAP Motion has its own limitations. We will investigate better sensor technologies to be implemented in a real-vehicle environment. We believe that iterative design and testing processes will contribute to theoretical advancements in multimodal display and control, as well as provide practical design guidelines, which will ultimately lead to better user experience and higher road safety.

References

Akyol, S., Canzler, U., Bengler, K., & Hahn, W. (2000). Gesture control for use in automobiles. In: Proceedings of the IAPR Conference on Machine Vision Applications (IAPR MVA 2000), Tokyo, Japan, pp 349–352

Alpern, M., & Minardo, K. (2003). Developing a car gesture interface for use as a secondary task. In: CHI'03 Extended Abstracts on Human Factors in Computing Systems, FL, USA pp 932–933 ACM. https://doi.org/10.1145/765891.766078

Bach, K. M., Jæger, M. G., Skov, M. B., & Thomassen, N. G. (2008). Evaluating driver attention and driving behaviour: Comparing controlled driving and simulated driving. In: Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, Swindon, UK, pp 193–201

Balakrishnan, R. (2004). "Beating" Fitts' law: virtual enhancements for pointing facilitation. International Journal of Human-Computer Studies 61(6): 857–874. https://doi.org/10.1016/j.ijhcs.2004.09.002

Basil, M. D. (1994). Multiple resource theory I: Application to television viewing. Communication Research 21(2): 177–207. https://doi.org/10.1177/009365094021002003

Bilius, L. B., & Vatavu, R. D. (2020). A multistudy investigation of drivers and passengers' gesture and voice input preferences for in-vehicle interactions. Journal of Intelligent Transportation Systems 25(2): 197–220. https://doi.org/10.1080/15472450.2020.1846127

Brewster, S. (1998a). The design of sonically-enhanced widgets. Interacting with Computers 11(2): 211–235. https://doi.org/10.1016/S0953-5438(98)00028-9

Brewster, S. (1998b). Sonically-enhanced drag and drop. In: Proceedings of the International Conference on Auditory Display, Glasgow, UK, pp 1–7

Burnett, G. E., Summerskill, S. J., & Porter, J. M. (2004). On-the-move destination entry for vehicle navigation systems: Unsafe by any means? Behaviour & Information Technology 23(4): 265–272. https://doi.org/10.1080/01449290410001669950

Caird, J. K., Johnston, K. A., Willness, C. R., Asbridge, M., & Steel, P. (2014). A meta-analysis of the effects of texting on driving. Accident Analysis & Prevention 71: 311–318. https://doi.org/10.1016/j.aap.2014.06.005

Cairnie, I. W. Ricketts, S. J. Mckenna And G. Mcallister, A. (2000). Using finger-pointing to operate secondary controls in automobiles. In: Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No.00TH8511) Dearborn, MI, USA, pp 550–555. https://doi.org/10.1109/IVS.2000.898405

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J. D., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., & Bucher, C. (2006). The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment. No. DOT-HS-810-593. United States. Department of Transportation. National Highway Traffic Safety Administration. https://rosap.ntl.bts.gov/view/dot/37370

Dingus, T. A., McGehee, D. V., Hulse, M. C., Jahns, S. K., Manakkal, N., Mollenhauer, M. A., & Fleischman, R. N. (1995). TravTek evaluation task C3-Camera car study. No. FHWA-RD-94-076. United States. Department of Transportation. Federal Highway Administration. https://vtechworks.lib.vt.edu/handle/10919/55071

Donmez, B., Boyle, L., and Lee, J. (2010). Differences in Off-Road Glances: Effects on Young Drivers' Performance. Journal of Transportation and Engineering 136(5): 403–409. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000068

Edwards, A. D. (1989). Soundtrack: An auditory interface for blind users. Human-Computer Interaction 4(1): 45–66. https://www.tandfonline.com/doi/abs/10.1207/s15327051hci0401_2

Elliott, D., Helsen, W. F., & Chua, R. (2001). A century later: Woodworth's (1899) two-component model of goal-directed aiming. Psychological Bulletin 127(3): 342–357. https://doi.org/10.1037/0033-2909.127.3.342

Gable, T. M., Raja, S. R., Samuels, D. P., & Walker, B. N. (2015). Exploring and evaluating the capabilities of Kinect v2 in a driving simulator environment. In: Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM. Nottingham, UK, pp 297–304. https://doi.org/10.1145/2799250.2799276

Gaver, W. W. (1989). The SonicFinder: An interface that uses auditory icons. Human-Computer Interaction, 4(1): 67–94. https://www.tandfonline.com/doi/abs/10.1207/s15327051hci0401_3

Green, P. (1999). Visual and task demands of driver information systems (No. UMTRI-98-16). The University of Michigan Transportation Research Institute.

Goulati, A., & Szostak, D. (2011, August). User experience in speech recognition of navigation devices: an assessment. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, Stockholm, Sweden, pp 517–520. https://doi.org/10.1145/2037373.2037451

Green, P. (2000, November). Crashes induced by driver information systems and what can be done to reduce them. No. 2000-01-C008. SAE Technical Paper, pp 27–36. SAE. https://www.sae.org/publications/technical-papers/content/2000-01-C008/

Grossman, T., & Balakrishnan, R. (2004). Pointing at trivariate targets in 3D environments. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp 447–454. ACM. https://doi.org/10.1145/985692.985749

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in Psychology 52: 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hassel, E. (2016). Touch screens in cars: Investigating touch gestures and audio feedback in the context of in-vehicle infotainment. Honors Thesis, Malmö University, Sweden. diva2:1482085

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. Human Factors 48(1): 196–205. https://doi.org/10.1518/001872006776412135

Hurwitz, J. B., & Wheatley, D. J. (2002,). Using driver performance measures to estimate workload. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting 46(22): pp 1804–1808. SAGE Publications. https://doi.org/10.1177/15419312020460220

Jeon, M., Gable, T. M., Davison, B. K., Nees, M., Wilson, J., & Walker, B. N. (2015). Menu navigation with in-vehicle technologies: Auditory menu cues improve dual task performance, preference, and workload. International Journal of Human-Computer Interaction 31(1): 1–16. https://doi.org/10.1080/10447318.2014.925774

Jahani, H., Alyamani, H. J., Kavakli, M., Dey, A., & Billinghurst, M. (2017). User evaluation of hand gestures for designing an intelligent in-vehicle interface. In: Maedche, A., vom Brocke, J., Hevner, A. (eds.) Designing the Digital Transformation. DESRIST 2017.

Lecture Notes in Computer Science(), 10243: 104–121. Springer, Cham. https://doi.org/10.1007/978-3-319-59144-5_7

Jiang, L., Xia, M., Liu, X., & Bai, F. (2020). Givs: fine-grained gesture control for mobile devices in driving environments. IEEE Access 8: 49229–49243. doi: 10.1109/ACCESS.2020.2971849

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. No. DOT HS 810 594. United States. Department of Transportation. National Highway Traffic Safety Administration. http://hdl.handle.net/10919/55090

Liu, Y.-C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveler information systems. Ergonomics, 44: 425–442. https://doi.org/10.1080/00140130010011369

Liu, Q., Ren, J. Qian, Z., Hua, M., (2017). Seated reach capabilities for ergonomic design and evalution with consideration of reach difficulties. Applied Ergonomics, 59(A): 357–363. https://doi.org/10.1016/j.apergo.2016.09.011

May, K. R., Gable, T. M., & Walker, B. N. (2014). A multimodal air gesture interface for in vehicle menu navigation. In: Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA. pp 1– 6. ACM. http://dx.doi.org/10.1145/2667239.2667280

McGuffin, M. J., & Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. ACM Transactions on Computer-Human Interaction (TOCHI), 12(4): 388–422. https://doi.org/10.1145/1121112.1121115

Mynatt, E. D., & Edwards, W. K. (1992). The Mercator environment: A nonvisual interface to X Windows and Unix workstations. GVU Tech Report GIT-GVU-92-05. Graphics, Visualization and Usability Center, College of Computing, Georgia Institute of Technology. Atlanta, GA, USA

National Highway Traffic Safety Administration. (2012). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. No. NHTSA-2010-0053. United States. Department of Transportation, National Highway Traffic Safety Administration. pp 1–117

Ohn-Bar, E., Tran, C., & Trivedi, M. (2012). Hand gesture-based visual user interface for infotainment. In: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, NH, USA, pp 111–115 ACM. https://doi.org/10.1145/2390256.2390274

Olson, R. L., Hanowski, R. J., Hickman, J. S., & Bocanegra, J. L. (2009). Driver distraction in commercial vehicle operations, No. FMCSA-RRR-09-042. United States. Department of Transportation. Federal Motor Carrier Safety Administration. https://doi.org/10.21949/1502647

Peng, Y., Boyle, L., and Hallmark, S., (2013). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. Accident Analysis and Prevention 50: 628–634. https://doi.org/10.1016/j.aap.2012.06.013

Pickering, C. A., Bumnham, K. J., & Richardson, M. J. (2007). A review of automotive human machine interface technologies and techniques to reduce driver distraction. In: Proceedings of the 2nd Institution of Engineering and Technology International Conference on System Safety, London, pp 223–228. doi: 10.1049/cp:20070468

Pfleging, B., Schneegass, S., & Schmidt, A. (2012). Multimodal interaction in the car: Combining speech and gestures on the steering wheel. In: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, NH, USA, pp 155–162. https://doi.org/10.1145/2390256.2390282

Rahman, A. S. M. M., Saboune, J., Saddik, A. El, & Ave, K. E. (2011). Motion-path based in car gesture control of the multimedia devices. In: Proceedings of the first ACM International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications, Miami, FL, USA, pp 69–75. https://doi.org/10.1145/2069000.2069013

Ranney, T. A., Mazzae, E., Garrott, R., & Goodman, M. J. (2000). NHTSA driver distraction research: Past, present, and future. No. 2001-06-0177. SAE Technical Paper. https://www.sae.org/publications/technical-papers/content/2001-06-0177/

Shakeri, G., Williamson, J. H. and Brewster, S. (2017) Novel multimodal feedback techniques for in-car mid-air gesture interaction. In: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, pp 84–93. ACM. https://doi.org/10.1145/3122986.3123011

Shneiderman, B. (1997). Direct manipulation for comprehensible, predictable and controllable user interfaces. In Designing the user interface (3rd ed., pp 33–39). Reading, MA: Addison-Wesley.

Society of Automotive Engineers (1988). Recommended Practice J287, Driver Hand Control Reach. Society of Automotive Engineers, Inc, Warrendale PA. https://www.sae.org/standards/content/j287_202211/

Society of Automotive Engineers (2007). Recommended Practice J287, Driver Hand

Control Reach. Society of Automotive Engineers, Inc, Warrendale PA. https://www.sae.org/standards/content/j287_202211/

Sodnik, J., Dicke, C., Tomažič, S., & Billinghurst, M. (2008). A user study of auditory versus visual interfaces for use while driving. International Journal of Human-Computer Studies 66(5): 318–332. https://doi.org/10.1016/j.ijhcs.2007.11.001

Srinivasan, R., & Jovanis, P. P. (1997). Effect of selected in-vehicle route guidance systems on driver reaction times. Human Factors 39(2): 200–215. https://doi.org/10.1518/001872097778543

Sterkenburg, J., Landry, S., & Jeon, M. (2019). Design and evaluation of auditory-supported air gesture controls in vehicles. Journal on Multimodal User Interfaces 13(2): 55–70. https://doi.org/10.1007/s12193-019-00298-8

Strayer, D. L., & Drew, F. A. (2004). Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers. Human Factors 46(4): 640–649. https://doi.org/10.1518/hfes.46.4.640.56806

Tabbarah, M. (2022). Novel in-vehicle gesture interactions: design and evaluation of auditory displays and menu generation interfaces. Unpublished Master's Thesis. Virginia Polytechnic Institute and State University.

Tijerina, L., Parmer, E., & Goodman, M. J. (1998). Driver workload assessment of route guidance system destination entry while driving: A test track study. In Proceedings of the 5th ITS World Congress, Seoul, Korea, pp 12–16

Tsimhoni, O., Smith, D., & Green, P. (2004). Address entry while driving: Speech recognition versus a touch-screen keyboard. Human Factors 46(4): 600–610. https://doi.org/10.1518/hfes.46.4.600.56813

Young, K. & Regan, M. (2007). Driver distraction: A review of the literature. In: I. J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J. D. Irwin (Eds.). Distracted driving. Sydney, NSW: Australasian College of Road Safety. pp 379–405.

Yu, X., Ren, J., Zhang, Q., Liu, Q., Liu, H. (2017). Modeling study of seated reach envelopes on spherical harmonics with consideration of the difficulty ratings. Applied Ergonomics 60: 220–230. https://doi.org/10.1016/j.apergo.2016.12.002

Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D. K., Dingler, T., & Jeon, M. (2013). Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. Human Factors 55(1): 157–182. https://doi.org/10.1177/0018720812450587

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science 3(2): 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D., & Seppelt, B. (2002). Interference with driving or in-vehicle task information: The effects of auditory versus visual delivery (Tech. Rep. No. AHFD-02-18/GM-02-3). Urbana-Champaign: University of Illinois at Urbana-Champaign, Aviation Human Factors Division.

Wierwille, W. W., & Tijerina, L. (1998). Modelling the relationship between driver in-vehicle visual demands and accident occurrence. Vision in Vehicles 6: 233–243.

Wu, S., Gable, T., May, K., Choi, Y. M., & Walker, B. N. (2016). Comparison of surface gestures and air gestures for in-vehicle menu navigation. Archives of Design Research, 29(4): 65–80. http://dx.doi.org/10.15187/adr.2016.11.29.4.65