2018

DESIGN AND EVALUATION OF AUDITORY-SUPPORTED AIR GESTURE CONTROLS IN VEHICLES

Jason Sterkenburg
Michigan Technological University, jtsterke@mtu.edu

DESIGN AND EVALUATION OF AUDITORY-SUPPORTED AIR GESTURE CONTROLS IN VEHICLES

By

Jason Sterkenburg

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Applied Cognitive Science and Human Factors.

Department of Cognitive and Learning Sciences

Dissertation Advisor: *Dr. Myounghoon Jeon*

Committee Member: *Dr. Keith Vertanen*

Committee Member: *Dr. Kevin Trewartha*

Committee Member: *Dr. Kelly Steelman*

Department Chair: *Dr. Susan Amato-Henderson*

Contents

Abstract

The number of visual distraction-caused crashes highlights a need for non-visual information displays in vehicles. Auditory-supported air gesture controls could fill that need. This dissertation covers four experiments that aim to explore the design auditory-supported air gesture system and examine its real-world influence on driving performance. The first three experiments compared different prototype gesture control designs as participants used the systems in a driving simulator. The fourth experiment sought to answer more basic questions about how auditory displays influence performance in target acquisition tasks. Results from experiment 1 offered optimism for the potential of auditory-supported displays for navigating simple menus by showing a decrease in off-road glance time compared to visual-only displays. Experiment 1 also showed a need to keep menu items small in number but large in size. Results from experiment 2 showed auditory-supported air gesture controls can result in safer driving performance relative to touchscreens, but at the cost of slight decrements in menu task performance. Results from experiment 3 showed that drivers can navigate through simple menu structures totally eyes-free, with no visual displays, even with less effort compared to visual displays and visual plus auditory displays. Experiment 4 showed that auditory displays convey information and allow for accurate target selection, but result in slower selections and relatively less accurate selections compared to displays with visual information, especially for more difficult target selections. Overall, the experimental data highlight potential for auditory-supported air gesture controls for increasing eyes-on-road time relative to visual displays both in touchscreens and air gesture controls. However, this benefit came at a slight cost to target selection performance as participants generally took longer to process auditory information in simple target acquisition tasks. Experimental results are discussed in the context of multiple resource theory and Fitts's law. Design guidelines and future work are also discussed.

Chapter 1

1 Introduction

The prevalence of touchscreens in vehicles has increased in recent years. 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 (Horrey & Wickens, 2007; Klauer et al., 2006; Olson, Hanowski, Hickman & Bocanegra, 2009; Wierwille & Tijerina, 1998; Dingus, et al., 2006) 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 (e.g., Microsoft Kinect). Some researchers have recently begun exploring 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). The purpose of this dissertation is to further develop and improve on these first efforts to create safer IVISs.

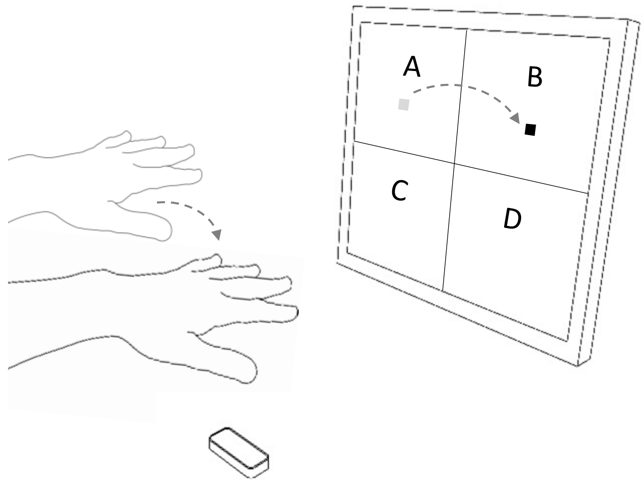

Figure 1.1 user moves from menu item A to menu item B with air gesture controls

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 (Figure 1.1). This is opposed to a symbolic system controlled via performance of dynamic gestures such as taps, swipes, or a type of sign language. Although such a system is possible and maybe even beneficial in some cases, my initial efforts were to develop a simple menu structure that represents the home page of typical in-vehicle touchscreen controls, with a selection of four to sixteen high-level menu items (e.g., audio, navigation, etc.).

In order 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. 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.

To evaluate the effectiveness of air gesture controls in reducing visual demands and improving driving safety, four experiments were conducted:

Experiment 1 aimed to determine what menu layout is more effective between a 2x2 grid with 4 four menu items and a 4x4 grid with 16 menu items, and also answer my questions about the impact of auditory displays on driving performance, secondary task performance, eye glance behavior, and driver workload.

Experiment 2 aimed to compare the performance of the best gesture control prototypes to emerge from Experiment 1 to touchscreen controls.

Experiment 3 addressed to emergent questions – 1) What is the impact of display control congruency and movement orientation? 2) What happens if we remove all visual information?

Experiment 4 was a basic experiment designed to answer more fundamental questions about how different auditory displays impact movement performance during simple target acquisitions.

1.1 Research Goals and Overview

The research objective of this dissertation is:

> *To improve understanding of how auditory displays and air gesture controls can be used effectively to enable safer driving through mitigation of visual and physical demands of in-vehicle information system use.*

In pursuit of this goal, several primary research questions were investigated, some of which build on results from previous experiments. The five primary research questions are:

a) How does the menu layout for air gesture controls influence driving performance, secondary task performance, eye glance behavior, and driver workload?
b) What effect do supplementary auditory displays have on driving performance, secondary task performance, eye glance behavior, driver workload, and preference?
c) What kind of auditory displays best facilitate goal-directed movement for simple target acquisitions?
d) How do these prototype systems compare to equivalent touchscreen systems in their influence on driving performance, secondary task performance, eye glance behavior, driver workload, and preference?
e) How does movement orientation influence driving performance, secondary task performance, eye glance behavior, driver workload, and system preferences?

To answer these research questions, the following research experiments were completed:

1.1.1 Experiment 1

Designed to answer questions (a) and (b), this experiment required participants to drive in a simulator while using four different air gesture control prototypes: a 2x2 grid with auditory feedback, a 2x2 grid without auditory feedback, a 4x4 grid with auditory feedback, and a 4x4 grid without auditory feedback. The 2x2 grids had a total of four large targets arranged in a square, while the 4x4 grids had sixteen total targets (Figure 1.2).

Prototypes with auditory feedback provided information about the current position of the cursor. For example, when a user holds their hand so the cursor is in the “A” target, then the system will say “A”. In addition, the auditory display provides non-speech tones as a confirmation of target selection. Results showed that using air gesture controls with 2x2 grids resulted in better driving performance, fewer off-road glances, better secondary task performance, and lower workload compared to 4x4 grid layouts. Results also showed that adding auditory displays had no impact on driving performance or secondary task performance, but did reduce the off-road glances and driver workload. More details can be found in Chapter 3.

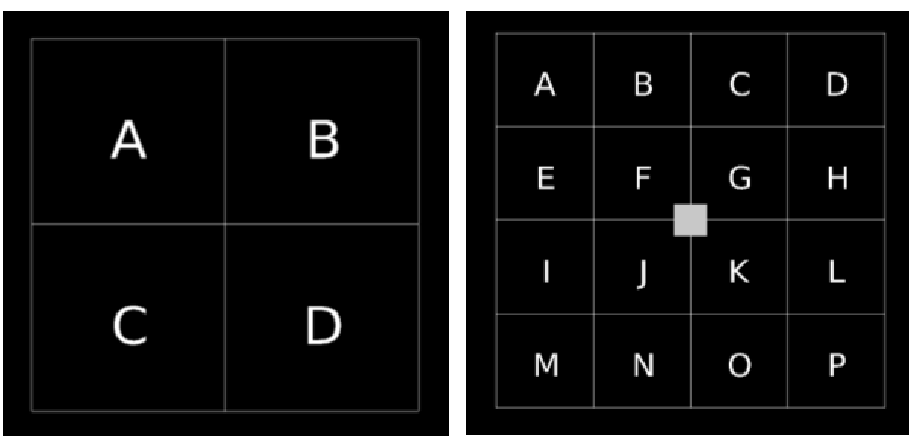

Figure 1.1 screen-capture of 2x2 grid (left) and 4x4 grid (right) used in Experiment 1

1.1.2 Experiment 2

Designed to learn more about questions (b) and (d), this experiment required participants to drive in a simulator while using 2x2 grids with and without audio for both air gesture controls and touchscreens. Results showed that driving performance and workload was equivalent between touchscreens and air gestures. Secondary task performance was worse using air gesture controls compared to touchscreens. However, the number of off-road glances was lower using air gesture controls with auditory displays. More details can be found in Chapter 4.

1.1.3 Experiment 3

Designed to answer questions (b) and (e), this proposed experiment required participants to drive in a simulator while using air gesture control systems with auditory displays with and without visual displays, and also using vertical as well as horizontal control orientations (Figure 1.3). Systems performed similarly with respect to driving performance, with participants spending slightly more time in the correct lane when using auditory-only controls. When examining secondary task performance, again, systems performed similarly but auditory-only controls led to slower task completion times. These results suggest a tradeoff between secondary task expediency and primary task performance. Workload was also lower for participants when using the auditory-only

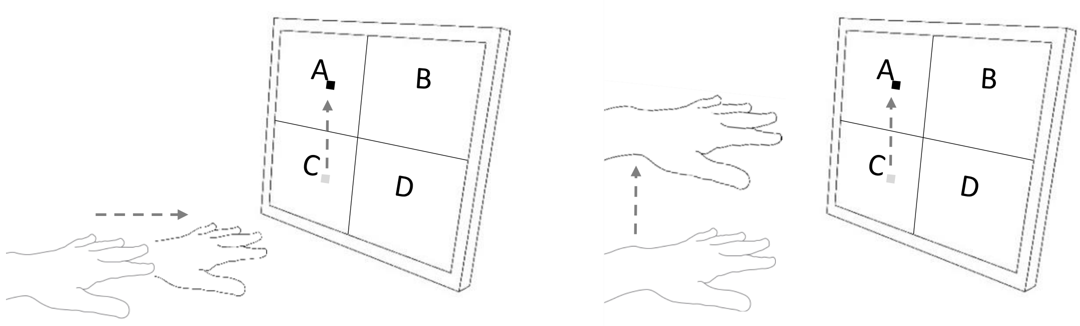

Figure 1.2 horizontal orientation (left) and vertical orientation (right)

display, a surprising finding. There was no difference between the vertical and horizontal control orientations, another surprising finding (or lack thereof). More details can be found in Chapter 5.

1.1.4 Experiment 4

Designed to answer (c), this experiment required participants to complete a serial tapping task, according to a traditional Fitts's task paradigm. However, instead of tapping on metal plates, or on a computer, participants completed the task in the air, using the LEAP Motion to measure hand position, and using combinations of visual and auditory displays to relay information about hand and target positions (Figure 1.4) Results showed that

Figure 1.4 traditional Fitts's task, image from (MacKenzie, 1992) (left), proposed setup for air gesture equivalent task (right)

mean selection times were slower for auditory-only displays (discrete and continuous) compared to displays with visual information available. Error (distance from selection point to target edge) was also lower for systems using only auditory displays relative to systems with visual displays while accuracy (percent of selections in target) was similar for all both auditory and visual displays. Both the speed and the accuracy degraded as difficulty increased, but degraded more dramatically with auditory-only displays. Throughput, a measure that combines both accuracy and speed was significantly higher (better) for conditions with visual stimuli compared to conditions without visual stimuli. More details can be found in Chapter 6.

1.2 Contributions

This proposed dissertation makes the following contributions to the areas of in-vehicle information systems, eyes-free gesture controls, auditory displays, and movement science.

1.2.1 In-vehicle Information Systems

In collaboration with lab members, I developed a novel air gesture IVIS that demonstrated ability to reduce off-road glances compared to touchscreens in a driving simulator. This research is among the first attempts to evaluate the viability of air gesture controls in vehicles for navigating simple menus while driving. While only a simple menu, this prototype successfully allowed for totally eyes-free interaction, improved time-in-lane, and lower driver workload compared to equivalent touchscreens, which has not been done to date, as far as I know. I have also defined early guidelines about appropriate target sizes, measured by index of difficulty (ID), to facilitate eyes-free selection of menu items. My research findings also highlighted performance decrements for in-air target acquisition associated with different areas within the reach envelope of drivers.

1.2.2 In-air Gesture Controls

Many of the findings that improve IVISs can also contribute more broadly to air gesture controls in other domains, especially virtual reality. These contributions include defining early guidelines for facilitating eyes-free target acquisition, and discovering uneven target acquisition performance within different areas within the reach envelope of air gesture control users. I also described human performance limits in a target acquisition task for systems that provide visual feedback, and different types of auditory feedback. This research provides researchers and designers with a basic expectation of the capabilities and limitations of eyes-free interaction performance for target acquisitions.

1.2.3 Auditory Displays

Through these experiments I have explored multiple sonification techniques, and measured their impact on driving and secondary task performance, as well as driver workload and preferences. By using different sonification techniques I can determine which ones minimize annoyance and overload, and identify techniques that best facilitate target acquisitions, using in-air gesture controls. In Experiment 4, I measured the throughput of movements made with visual information, and with auditory information, to more precisely define the influence of different auditory displays on target acquisitions. This will help auditory display researchers to know whether adding an auditory display will improve target acquisition performance for both visual and eyes-free interactions for a variety of movement difficulties. Furthermore, this research demonstrated the difference between adding a continuous vs a discrete auditory display. Results from Experiment 4, should also be applicable to many other scenarios involving the use of auditory displays in target acquisitions.

1.2.4 Movement Science

To my knowledge, there have been no studies done comparing movement performance in target acquisition tasks using auditory displays as a means of conveying information about relative position of user hand and target position. It may be of interest to some researchers in the movement science community to learn about the influence of continuous and discrete auditory displays on throughput for target acquisitions of a variety of movement difficulties.

1.3 Dissertation Outline

Chapter 2 summarizes relevant background work in multitasking while driving, eye glances and driving, auditory displays, and air gesture controls.

Chapter 3 presents Experiment 1, which investigated the impacts of grid layout and auditory displays on driving performance, secondary task performance, eye glance behavior, workload and preferences.

Chapter 4 presents Experiment 2, which used the same measures as Experiment 1, but compared air gesture controls to an equivalent touchscreen system.

Chapter 5 presents Experiment 3, which measured the impact of auditory displays, without corresponding visual displays, on driving performance, secondary task performance, eye glance behavior, and driver workload. This experiment also measured the impact of two movement orientations (vertical vs. horizontal) on those same measures.

Chapter 6 summarizes relevant background work in movement science, and presents Experiment 4 which measured the influence of combinations of visual and auditory displays on movement performance for target acquisitions using air gesture controls.

Chapter 7 draws conclusions, summarizes limitations, and suggests possible future work related to this dissertation.

Chapter 2

2 Literature Review

2.1 Related Work

This chapter presents work on multi-tasking in vehicles, auditory displays, and target acquisition. I begin by introducing the theoretical factors that influence drivers' abilities to multitask while driving, and support theoretical expectation by citing experimental results and naturalistic driving data. Next, I present the relationship between eye glance behavior and driving performance and summarize how auditory displays have been used in similar contexts to facilitate eyes-free device use. Finally, I present literature from movement science that can help me understand how well people will be able to acquire targets, a key task in menu navigation, using an eyes-free air gesture control system.

2.2 Multi-tasking in Vehicles

In-vehicle information systems (IVISs), like navigation devices, mobile phones, and radios require driver input to be used. When a driver wants to use an IVIS, he/she must balance the demands of the driving task with the demands of using the IVIS, often switching attention between the two. Multiple Resource Theory (Wickens, 2002) models how the demands of each task influence the performance when task-switching between multiple tasks. It suggests that while multi-tasking, performance on two or more tasks is dependent on their overlap in demand for resources. If two tasks share demands for the same finite resources then performance on one, or both tasks will suffer. Since driving and IVIS use are primarily visual-manual tasks, multiple resource theory predicts that driving performance may be degraded when drivers attempt to use IVISs, as long as those IVISs require visual-manual resources to use.

Still the question remains, how do we determine if two tasks are using the same resources? We need to answer this question to understand more deeply which tasks will interfere with the performance of which tasks. A deeper look at Multiple Resource Theory (MRT) shows that Wickens (1984) described task demands on three separate dimensions (Figure 2.1):

1a) Perceptual modality – comprised of visual and auditory subcategories. Describes what information channel is being used for a task.

1b) Visual channel – comprised of ambient or focal subcategories. Describes whether information is in focal visual area or in the periphery.

2) Processing format – comprised of spatial or verbal subcategories. Describes whether language is being processed by reading or listening.

3) Information processing stage – comprised of perception, cognition, response. Describes the three stages of information processing.

Every task we engage in throughout the day lies somewhere on each of these independent dimensions and each subcategory represents a unique pool of resources. If we examine a task, we can define its characteristics within these three dimensions (Figure 2.1). The degree to which another task overlaps for those resources will predict how much multitasking performance will be degraded. For example, lane keeping – the task of staying in correct lane while driving – requires the visual channel and ambient visual attention, and spatial processing, and requires all stages of information processing. Using an in-vehicle information system, requires focal visual attention and both verbal and spatial processing in the perception and cognition stages. From this analysis we can predict that drivers will encounter some difficulty as a result of the common demand for visual attention in the perceptual stage. Even though the overlap is not total, focal visual attention is necessary for both verbal and spatial processing for the IVIS use. The lane keeping task also requires spatial processing and visual attention. Even though the visual attention required for lane keeping is only ambient, the two types of visual resources are more similar, and therefore more prone to competition, than any visual resource is compared to auditory resources. Wickens' model (Figure 2.1) visualizes the three dimensions on a cube. In this model, each block represents a unique resource type. When multiple tasks require use from the same block of resources, conflict arises that degrades multitasking performance (Wickens, 1984). Of course, driving tasks are not limited to these subtasks. Drivers engage in a host of other activities which require focal visual attention, such as reading traffic signs, tracking movement of other vehicles and pedestrians, or listening to instructions from a navigation device.

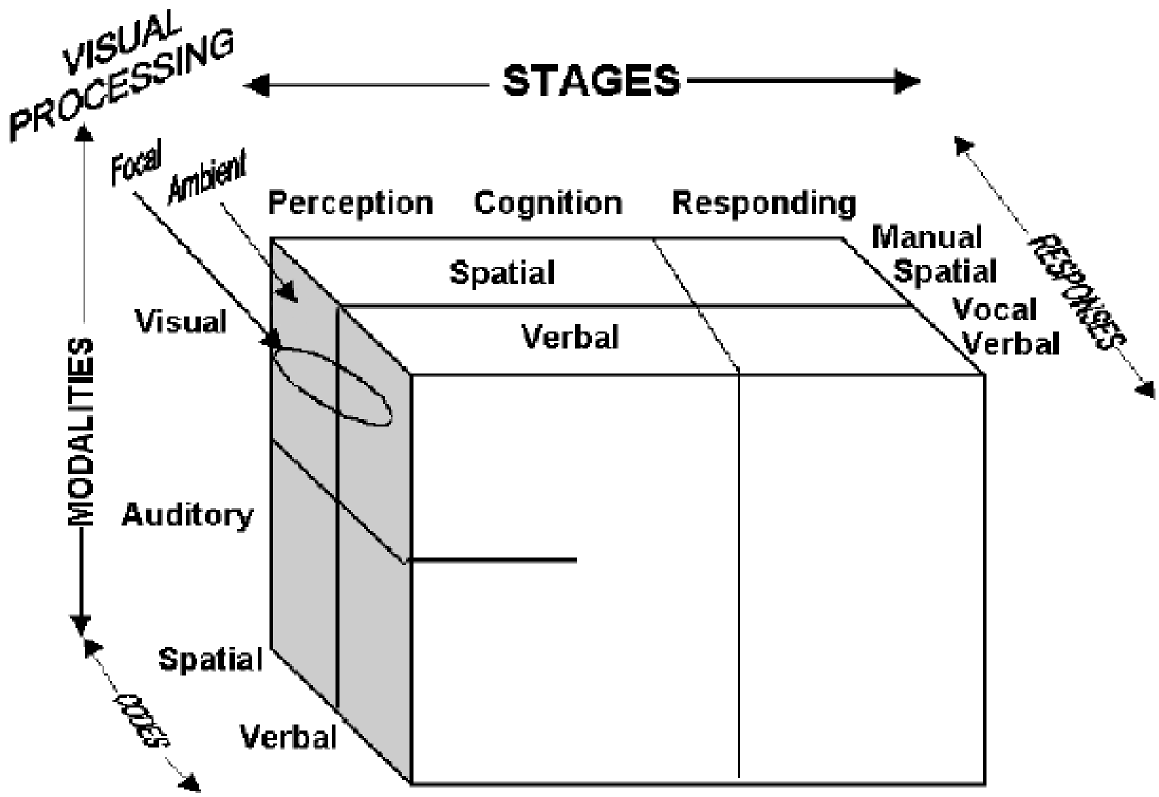

Figure 2.1 Multiple Resource Theory model (Wickens, 1984; image from Wickens 2002; copyright permission granted). Shows the three stages of processing and each of the subcategories used to categorize tasks according to their resource demands.

In his later work, Wickens (2002) also stated that task difficulty may impact multi-tasking performance. Difficult tasks demand relatively more resources and sap the remaining resources available to allocate to other tasks, even if both tasks do not use the same resources. This aspect of MRT explains why drivers can struggle to complete tasks that do not overlap for the same resources. For example, a driver who is driving on a straight empty highway may be able to carry a conversation with a passenger but the same driver may be unable to hold a conversation while driving in heavy traffic or at an unfamiliar intersection.

One of the strategies people use to manage the demands of multitasking is task-switching – moving their attention back and forth between multiple tasks. This strategy is necessitated when multiple tasks require focal visual attention because focal visual attention cannot be split between two locations (Hoffman & Subramaniam, 1995). For this reason, focal visual attention is often used as a proxy metric of attention (Granka, Joachims & Gay, 2004; Nielsen & Pernice, 2010; Goldberg, Stimson, Lewenstein, Scott & Wichansky, 2002). However, attention is not confined to vision. For example, a driver may look at the road, but really be focused on a conversation they are having with a passenger. In this case, the driver is looking at the road, but cognitive resources are being funneled to the conversation. Competition leading to diversion of attention away from driving to secondary tasks that results in degraded driving is called *driver distraction* (Young & Regan, 2007). Under this definition of driver distraction, IVIS use is a driver distraction, and so is talking to a passenger. 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. The utility of multiple resource theory in this pursuit is that MRT can describe and predict when conflicts for finite resources will give rise to distractions that will degrade driving performance.

2.2.1 Impacts of driver distraction

Since we just determined that IVIS use is a type of driver distraction, this section introduces literature on the influences of different types of driver distraction on driving performance.

2.2.1.1 Mobile devices

Much of the research in driver distraction has been related to the use of cell phones in vehicles. Texting, holding conversations, dialing phone numbers, and using route guidance apps have all been investigated. In this section I present a short summary of the research conducted on in-vehicle information systems.

Research has shown that engaging in secondary activities such as texting (Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009) and talking on a cell phone (Strayer, & Johnston, 2001; Horrey, & Wickens, 2006) degrade driving performance. 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).

The takeaway lesson from this research is that mobile devices use while driving leads to increased crash risk primarily by increasing reaction time to changes in the environment and not degraded lane keeping ability.

2.2.1.2 Infotainment systems

Cell phones do not present the only risk of in-vehicle information system distraction. Infotainment systems also require visual demands. Below I introduce a brief summary of research on the distracting effects of infotainment systems.

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. Younger adults (under 35) had less difficulty balancing the data entry with driving than older adults (over 55), who took twice as long, on average, to complete the destination entry tasks. Tijerna and colleagues also investigated the effects of menu structure on driving performance and eye glance behavior. They used a short list menu (3 items visible) navigated by a knob, a longer list menu (11 – 13 items visible) navigated using an arrow, or a keyboard layout navigated with a joystick. Their results showed that short list menu structures led to shorter task completion times, fewer off-road glances, and fewer lane departures. 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).

It is noteworthy that each of the results from the driving research literature aligns very well with the predictions from Multiple Resource Theory (MRT). For example, we would expect touchscreen controls to lead to degraded multitasking performance for drivers because they both require focal vision (at least to drive safely). Voice controls, on the other hand, do not require the same visual resources and therefore, we should expect relatively better multitasking performance.

2.2.2 Eye glances and driving

The driving literature clearly points to conflict for visual attention as one of the major causes of distraction-related crashes. We know that off-road glances are bad. However, not all off-road glances are equal in their impact on driving performance. According to 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. Long glances away from the road – 2 seconds or more – increase near-crash/crash risk by at least two times normal driving (Klauer, Dingus, Neale, Sudweeks & Ramsey, 2006).

To improve driving safety, the National Highway Traffic Safety Administration (NHTSA) developed guidelines for IVIS design that suggest limits for permissible visual demands of IVIS use (National Highway Traffic Safety Administration, 2012):

(1) Driver should be able to complete tasks while driving with glances away from the road of 2 seconds or less

(2) Cumulative eyes-off-road time should not exceed 12 seconds for a single task

The Alliance of Automobile Manufacturers (AAM) also produced voluntary guidelines for system designs (Boyle et al., 2013). Their principle 2.1 addresses distraction stating, "Systems with visual displays should be designed such that the driver can complete the desired task with sequential glances that are brief enough not to adversely affect driving." They state that there are two methods for verifying adherence to those guidelines:

(1) The 85^{th} percentile of all glance durations should not exceed 2 seconds

(2) The number of lane departures should not exceed those of a reference task, such as tuning a radio

These guidelines and principles informed the design and analysis of Experiments 1, 2, and 3 and will inform future iterations of the prototype design and future evaluations of the prototype effectiveness.

While the automotive standards and guidelines focus on the role of focal visual attention on driving performance, another important aspect of visual attention is peripheral visual attention. Research has shown that useful field of view (UFOV) is a better predictor of driving performance than performance on a battery of tasks including a visual acuity task, brake reaction time task, split-attention task, and sign recognition task (Myers, Ball, Kalina, Roth, & Goode, 2000). Useful field of view includes is the area from which people can gather information at a glance without moving their head or eyes (Ball, Wadley, & Edwards, 2002). The observation that UFOV was a better predictor of driving performance than brake reaction time is interesting as apparently contradicts Horry and Wickens' finding (2006) that impacts of distraction were mediated primarily by delays in reaction time. The observation from Myers et al. that UFOV is a better predictor of driving performance than performance on a reaction time task likely arose because there were no distractions during their driving scenario. Therefore, it is still the case that distraction-related driving performance degradation is mediated by increased reaction times, but it is also true that during un-distracted driving, UFOV can be a better predictor of driving performance. The effectiveness of UFOV as a predictor makes theoretical sense because peripheral vision is also necessary for important driving tasks such as lane keeping, monitoring positions of other agents on the road, locating and identifying posted signs, and detecting changes in traffic lights. Based on the experimental observations and theoretical principles, it would be fair to consider individual differences in UFOV as an important factor in driving performance.

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.

2.3 Auditory Displays

Why use auditory displays?

There are a few basic factors to consider when establishing the case for using auditory displays in vehicles. First, the human auditory system is tuned to detect patterns in sound over time (Bregman, 1990; Kramer et al., 1999). Secondly, multiple resource theory predicts a competition for similar resources that could result in performance degradation and distraction. This conflict for visual attention is recognized by the in-vehicle display guidelines that require in-vehicle controls be usable with only very short visual glances. Past meta-analytic studies have also demonstrated that auditory displays or multimodal displays that provide visual and auditory information outperform visual-only displays in vehicles (Wickens & Seppelt, 2002; Liu, 2001). This rules out visual displays, but what about tactile or haptic displays? Well, one of the major benefits of auditory displays is that the auditory system can receive information from any direction, at any time, whereas a tactile display requires contact with a vibrating surface. Implementation of a haptic system would either need to accept the risk that a driver would not be contacting the display surface. This practical consideration means that auditory displays are more suitable for conveying information in vehicles.

Auditory displays in target selection tasks

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).

Previous research comparing target selection task performance across visual, auditory, and tactile presentation modalities has shown that audio feedback and tactile displays resulted in similar numbers of overall successful task selections completed but visual displays resulted in comparatively more successful task completions (Charoenchaimonkon, Janecek, Dailey, & Suchato, 2010). This research also showed that participants were more accurate when using the visual displays compared to auditory or tactile displays, which performed similarly. Other researchers (e.g., Akamastu, MacKenzie, & Hasbrouq, 1995) have also shown that tactile feedback can be processed more quickly, and result in shorter times between displaying feedback (cursor in target area triggers feedback, but do not impact the overall time it took from the start to end of a selection task. Shakeri, Williamson, and Brewster (2017) showed that auditory displays led to relatively better performance using an in-vehicle gesture system. A meta-analysis of the impact of multimodal displays on user performance showed that adding auditory or tactile displays improves reaction times to stimuli but does not reduce error rates for target acquisition tasks (Burke et al., 2006). Burke and colleagues also noted from their meta-analysis that tactile-visual displays generally led to better user performance for systems that were used in high workload conditions or when multitasking. However, these effects were only consistently observed for target acquisition tasks and were mediated by task type. It is important to consider the type of task associated with IVIS use. While target acquisitions are included, searching is also an important task. Efficient searching requires much more information than simple feedback needed to target acquisition. Since the auditory system can process information at a higher bandwidth, it is a more appropriate modality to facilitate searching. On balance, the existing literature suggests that tactile displays may or may not confer a small benefit over auditory displays leading to faster reaction-times to feedback for target acquisition tasks. Yet auditory displays are still necessary to facilitate searching that users are required to do to use an in-vehicle menu. Therefore, in an effort to develop a simple working system, it appears reasonable to develop an auditory-supported system and forgo the minimal additional benefits of tactile displays.

In-vehicle auditory displays

In-vehicle controls, if supported with appropriate auditory feedback, may limit visual demands and allow drivers to navigate menus and controls without looking away from the road. Sonification – the use of non-speech audio to convey information – (Kramer, 1993) can provide information about the position of the hand, and the gap between the current position and target position. However, existing guidelines (e.g., Driver Focus-Telematics Working Group, 2002) provide little help in the design of in-vehicle auditory displays, leaving designers with many unanswered questions about best practices for in-vehicle auditory display design.

Despite the apparent lack of official dictation of best practices, it is possible to glean some basic guidelines about in-vehicle auditory display design from auditory display literature. Nees and Walker (2011) reviewed the auditory display literature and described three basic axioms of auditory display design thus:

Detectability – use sounds that people can hear.

Discriminability – when sounds are used to represent distinct system states, use sounds that people can perceive as being different.

Identification – use sounds for which people can identify the intended meaning.

Now we can examine each of these basic axioms more deeply to determine how to consider each aspect in an auditory display.

Detectability

People are only sensitive to sounds within a certain range of frequency. Keeping sounds between 100 -10,000 Hz will help, but maximal sensitivity is between 2000 Hz and 5000 Hz (Gelfand, 2009). This fact cannot be easily translated into a universal design because of the influence of auditory masking from other sound sources inside and outside the vehicle. It is impossible to give an exact guide because there are a number of variables than impact the frequency of background cabin and road noise. For the purposes of the following experiments, the frequency can be reasonably ignored because of the aforementioned reasons, but it is also a matter more relevant to real-world implementation rather and is not immediately relevant to the more fundamental questions that are the subject of this dissertation.

Discriminability

People should be able to tell two distinct signals apart. Sound parameters that influence discriminability include: pitch (Stevens, Volkmann, & Newmann, 1937), loudness

(Stevens, 1936), tempo (Boltz, 1998), duration (Jeon & Fricke, 1997), background noise and signal similarity (Aiken & Lau, 1966), and time lapsed between signals (Aiken, Shennum, & Thomas, 1974). Again, because the in-vehicle environment is complex, it is difficult to prescribe specific parameters to follow for each aspect of a sound. The best practice as described by Neese and Walker (2011) is to avoid thresholds of discrimination as much as possible.

Identification

People should be able to associate the appropriate meaning with each sound. Ability to identify sound meanings is limited to a small set when using abstract sounds (Watson & Kidd, 1994). More ecologically meaningful sounds can be easier to identify (Bonebright & Nees, 2007; McKeown & Isherwood, 2007). Identification is a common problem in modern vehicles because of the amount of abstract sounds used in in-vehicle information systems.

The use of speech can facilitate both discriminability and identification. Discriminability is easily achieved due to the heightened sensitivity to even small differences in speech patterns. Identification is also more easily achieved due to the pre-existing mappings of meanings to speech sounds.

Research has shown that detectability, discriminability, and ability to identify sounds become more difficult as the number of concurrent sounds increases, especially if the sounds are similar (Bonebright, Nees, Connerley, & McCain, 2001; Walker & Lindsay, 2006). Another common concern associated with auditory displays is annoyance

(Edworthy, 1998; Kramer, 1994). Too many sounds can over-saturate the vehicle cabin and overwhelm drivers. Although not problems unique to the auditory modality, false alarms and misses can contribute to the annoyance of listeners. These are among the factors that designers should consider when developing an auditory display for in-vehicle use.

Overall, these basic axioms provide a general guideline that suggests, restricting the frequency range to that most sensitive to the human ear, distinguishing sounds as much as possible on each of the sound parameters to facilitate discrimination and avoiding confusion, and using ecologically representative sounds that cue listeners to the meaning associated with the sound. Speech appears to offer an easy path to discrimination and identification, which makes it a potential design element to include within an in-vehicle information system.

2.4 Air Gestures in Vehicles

Why use air gesture controls?

If drivers are required to move their hands over the surface of the screen to search and navigate through the menu, it will require a lot of hand-on-touchscreen time. 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. Overall, research has shown that tactile feedback overall offers no benefits over auditory displays in terms of task completion times, and an auditory-only touchscreen display would require moving touchscreens into a more reachable area of the vehicle. Meanwhile, gesture sensors can record movement data within a wide range of space, allowing for less physically demanding reaching movements for drivers.

In-vehicle air gesture controls

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. Some 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 and/or its impact on driving (Akyol et al., 2000; Cairnie et al., 2000; Rahman et al., 2011). In this dissertation, I attempt to both develop and evaluate a working prototype air gesture control system through an iterative design process.

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 when using the air-gesture control system. Despite mixed results, eye glance behavior was still within NHTSA guidelines (National Highway Traffic Safety Administration, 2012). This study demonstrates the feasibility of air gesture controls in vehicles. I was interested in furthering this line of research and developing a system that may improve driving performance and reduce off-road eye glances relative to touchscreens.

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.

The potential advantages of an air-based gesture control system over a touch-based system remains an open question. Auditory-supported touch interfaces have been demonstrated to be helpful in navigating through long song lists using systems such as the Earpod (Zhao, Dragicevic, Chignell, Balakrishnan, & Baudisch, 2007) and have been widely used by blind people to facilitate touchscreen use, utilizing the slide rule technique (Kane, Bagham, & Wobbrock, 2008). 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 improve upon 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.

2.5 Menu structure

One more important line of questioning surrounds the menu structure. The experiments detailed in this dissertation investigated very simple grid menu structures, resembling the home page of many in-vehicle infotainment systems. However, infotainment systems generally have many different menu structures at lower levels in the menu hierarchy (e.g., lists for audio tracks). It remains an open question how other menu structures can be designed to be used safely and eyes-free. What is the best way to design a gesture controlled list menu for eyes-free navigation? What is the impact of breadth versus depth in arranging menu items? Previous research has shown that there is an optimal tradeoff when searching through a 64-item hierarchical menu (Miller, 1981) with 2 levels of 8 menu items outperformed other displays of 1 level of 64 items, 3 levels with 4 items, 6 levels of 2 items. Miller also suggests that, if menus require more items, that menus expand in breadth rather than depth. Related studies have shown similar results (Zaphiris, 2000; Snowberry, Parkinson, & Sisson, 1983; Jacko, Salvendy, & Koubek, 1996), all generally pointing in the direction that breadth outperforms depth with respect to task completion times, error rates, and perceived complexity. However, all of those studies were investigating menus designed to be used on computers where users can devote 100% of attention to the menu. Drivers, in contrast to computer users, cannot devote all of their attention to the searching for and selecting a target. Research by Manes and Green (1997) showed a 26% increase in off-road eye glances and a 14% increase in lane departures when drivers used broad menus as compared to deep menus. So, in the context of driving, the optimal balance of depth/breadth shifts towards depth because of visual demand required to search through broad menus is higher. Hick (1952) and Hyman (1953) showed in their research that it takes longer to react to a visual stimulus as more information is presented. The Hick-Hyman law says that there is a logarithmic relationship between the number of items presented and the time to identify a target among the items. The common explanation for the logarithmic relationship is that as people visually scan for their target by eliminating half of the items, and then half again, until reaching the single target item. However, this explanation and model only works when people can anticipate the approximate location of the target (Cockburn & Gutwin, 2009). If approximate knowledge of the position is not known, a linear relationship describes the relationship between the number of items and response time (Cockburn & Gutwin, 2009). The Hick-Hyman law explains why broader menu structures take more visual attention and could therefore pose a greater threat to drivers than deeper menu structures in theory. One interesting study, conducted by Burnett, Lawson, Donkor, and Kuriyagawa (2013), demonstrated that broader menus were superior to deep menus in task completion times both with and without visual occlusions (in accordance with ISO standards: Klauer et al., 2006) when the menu information was structured alphabetically, but there was no difference between broad versus deep menus when the information was unstructured, leading the authors to the conclusion that broader menus were generally preferred. The experiment from Burnett et al. (2013) was conducted in a driving simulator but did not require the participants to drive, which calls into question the external validity of their results and suitability of the conclusions. Overall, the existing literature shows that broader menu structures facilitate faster, less error prone, and easier target selections. Even when vision is limited to 1.5 second windows, as is done in standard occlusion studies, broader menus still outperform deep menus. However, in real world driving environments, driving is degraded when using broader menus relative to deeper menu structures, leading to the conclusion that in-vehicle menus should be tend to be deep rather than broad. More research will need to be done to see if these trends hold up when considering factors such as input method (e.g., gesture controls) or visual display size.

2.6 Target acquisition

Using air gesture control systems fundamentally requires simple target acquisitions, similar to touchscreen use. That is, users are moving their hands through space, towards a target with the intention of selecting that target. It is important to learn what target sizes are feasible for in-vehicle gesture controls and what the impact of adjusting target sizes is on task completion times and accuracy.

Paul Fitts's seminal work, in which he described the relationship among movement difficulty, movement distance, and target size (Fitts, 1954; Fitts & Peterson 1964) allows us to predict what targets will be more difficult to select and provides a means to compare the difficulty of target acquisition tasks. The modified Shannon formulation of Fitts's original formula is the most frequently used in HCI (1) (MacKenzie, 1992).

$$ID = log_2\left(\frac{A}{W} + 1\right), \quad (1)$$

Here, A is amplitude, or distance from the start of the movement to the target and W is the target width. Index of difficulty (ID) is logarithmically proportional to the ratio of distance to target width. In other words, as the distance between the starting point and the target increases and/or the target width decreases, the difficulty of the movement (ID) increases. Fitts also showed that ID has a positive correlation with movement time and error rates (Fitts, 1954). From this I can predict that menu layouts with smaller targets will be more difficult to use, i.e., movements will take longer, and have increased error rates compared to menus with larger targets.

Research has shown that movements along the z-plane (forward/backward) are slower and more error prone compared to movements along the x (left/right) and y axes (up/down) (Grossman & Balakrishnan, 2004; Cockburn et al., 2011). Cockburn points out that their results may be influenced by the inability for their participants to easily perceive depth on their 2D visual display. However, Grossman and Balakrishnan found the same result using a volumetric display, which do not have the same limitations (Cockburn et al., 2011).

Chapter 3

3 Experiment 1 – Investigating Impacts of Menu Layout and Auditory Displays

3.1 Introduction

We developed four prototype systems: a 2x2 grid with auditory feedback, a 2x2 grid without auditory feedback, a 4x4 grid with auditory feedback, and a 4x4 grid without auditory feedback (Figure 3.1). Each of these prototype systems was created to investigate the influence of grid layout and auditory feedback on vehicle speed, lateral vehicle control, frequency of off-road glances, secondary task performance, and driver workload.

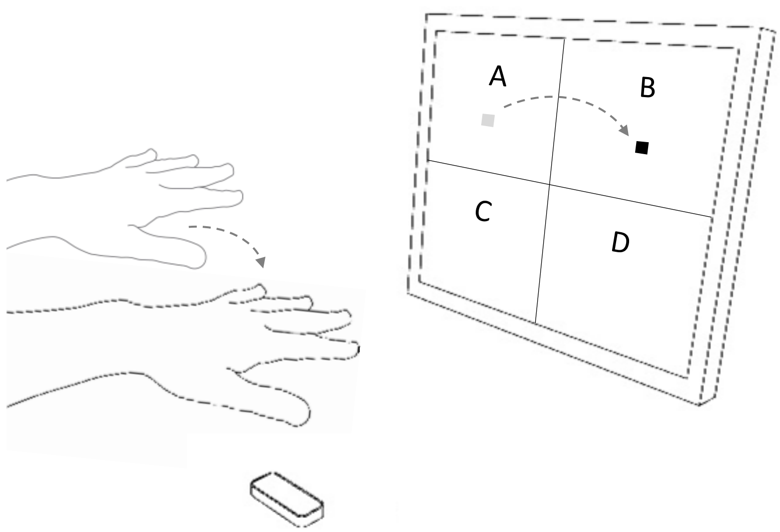

Figure 3.1 using air gestures to move from target A to target B in a 2x2 grid

3.2 Design Process

The purpose of the development air gesture control system prototypes was to evaluate their potential. From the beginning of the project, the design process was expected to be iterative, with learnings and questions from each experiment informing potentially improved prototype designs for the next experiment.

As a first test, we decided to develop several simple menus in order to benchmark their relative performance. The menus, while not representative of the full depth seen in many modern in-vehicle menus, provide a chance to observe differences between different designs, while also being much simpler to create. The expectation is that observed differences with a simple menu may be exaggerated in a more complex menu, so we should be careful not to over-generalize from observations made with drivers using these simple menus.

In this experiment there were 2x2 grids, with only four square targets, 5x5 inches across and 4x4 grids, with sixteen 2.5x2.5 inch targets. The 2x2 grids and 4x4 grids were chosen because they represented large size differences with 2x2 grid targets being twice the width and four times the area of the 4x4 grid targets. This allows for a clear method to determine the influence of target size on the dependent measures.

The addition of auditory displays was to investigate the influence of an auditory display in combination with an air gesture control on the dependent measures. The auditory display was a speech readout of the target name. This was done to allow for the greatest learnability of the system. With speech, as opposed to more abstract non-speech sounds, drivers have to dedicate less mental effort to learning and remembering the associated meaning. In reality, it may not always be realistic to represent all menu item names with a speech sound. In fact, a combination of speech and abstract sounds may ultimately be the best because it could allow for expert users to quickly search through complex menus while novice users can wait for the speech readout to provide more information. For the purposes of the first experiment the auditory display was intended to be highly learnable to avoid the need to go through a lengthy learning process. However, this does mean that use of gesture controls with auditory displays could be a little slower than its potential with non-speech sounds.

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 also so they have confidence that the system is responding to them (cursor and selection highlight).

The auditory display was made to mirror the information conveyed by the visual display as much as practically possible. There was a speech readout of each target name to mirror the visual text display. There was a raindrop sound corresponding with menu items selections for which the visual displays show a colored highlighted box. The only visual component for which there was no corresponding auditory element was the cursor position. While the auditory display allowed users to tell which menu item they had selected, it did not give them any more fine grain information about their position which the cursor could with the visual display. It is potentially possible to convey this information in an auditory display through an auditory display by using non-speech sounds and modulating non-speech sounds according to distances between the cursor and absolute positions within the menu. However, this would have potentially been more complicated system for drivers to learn and would have added more time to pre-drive training, and it could also be very distracting and noisy. It is possible that a continuously adapting auditory display could be more of a detriment than a benefit to drivers. Yet, there is still potential in this type of auditory display if it could be executed in a way that it does not overwhelm drivers.

The selection gesture, i.e., the gesture that drivers make in order 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 in order 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.

The movement plan was a horizontal plane, following the metaphor of a mouse movement with a computer. This is not necessarily the way the system had to work since air gesture controls are not limited to two-dimensional space. We chose this orientation because we assumed it would be less physically demanding than movements on the vertical plane. This decision was later a topic of research in Experiment 3.

3.3 Hypotheses

H1: Fewer/larger target sizes (2x2 grids) will reduce the secondary task difficulty and result in fewer lane departures and fewer off-road glances of all durations compared to more/smaller target sizes (4x4 grids).

H2: Auditory feedback will decrease secondary task difficulty and result in fewer lane departures, fewer off-road glances, and faster, more accurate selections while using the prototypes compared to conditions without auditory feedback.

3.4 Methods

3.4.1 Participants

A total of 23 participants, 14 males and 9 females, were recruited from the undergraduate psychology student pool at Michigan Technological University (Table 3.1).

Table 3.1 Participant demographics

	Age	**Experience (yrs)**	**Miles/yr**
Mean	19.9	4.13	8589
SD	1.53	1.55	6838

3.4.2 Gesture Control Prototypes

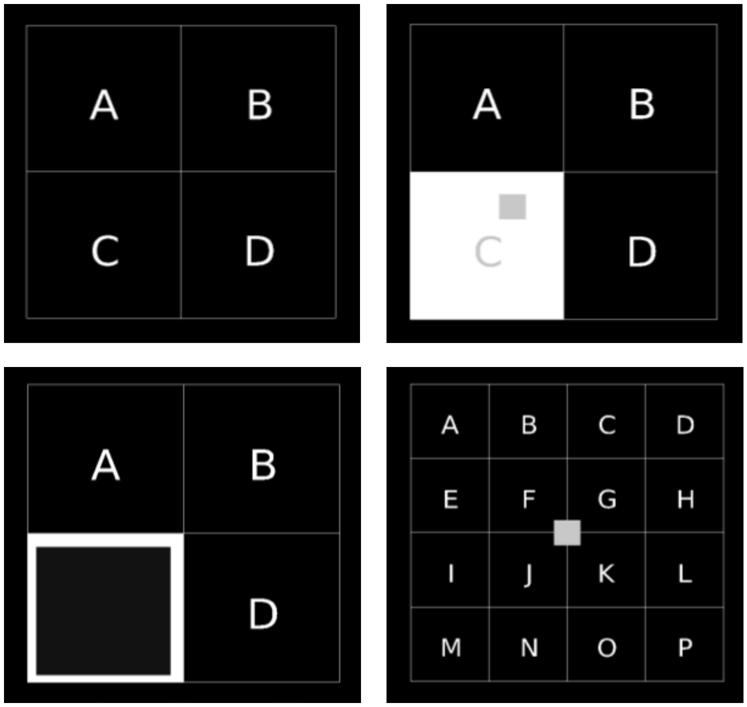

Figure 3.2 2x2 grid (Top Left), 2x2 grid with visualization of hand position and highlighting box C (Top Right), 2x2 grid showing visualization of a selection (Bottom Left), and Graphical display of 4x4 grid with hand position (Bottom Right).

The in-vehicle air gesture interface was comprised of two major components. A LEAP Motion, an infrared sensor designed to recognize hand features, was used to detect the hand position of the driver. The LEAP Motion sends data to Pure Data, an open-source, real-time graphical programming environment for audio and visual processing. Using the Pure Data patch we generated audio and visual displays incorporating the LEAP Motion data. Visual displays for all four air gesture prototypes were comprised of a number of target boxes arranged in a grid (Figure 3.2) – the 2x2 grids contained a total of four larger targets (5 in. by 5 in.) and the 4x4 grids contained 16 smaller targets (2.5 in. by 2.5 in.). There were two versions for each grid layout, one with auditory feedback and one without. Each target box contained a letter. As the user holds his/her hand over the LEAP Motion, the monitor displays a square cursor representing the position of the user's hand within the grid. If the center of the user's hand is within one of the boxes, that box is highlighted (Figure 3.2). For prototypes that have auditory feedback, the same action will play a wave file containing a text-to-speech readout for the name of the target that is currently highlighted. Navigation and target selection is dependent on the number of fingers visible to the LEAP Motion sensor. If the system detects five fingers, then it will select the target, which is highlighted at that moment. For the prototypes that have auditory feedback, a selection action is followed by a confirmatory earcon, which contains two "raindrop" tones, the first low followed immediately by a second higher frequency note. This is intended to provide an indication of selection.

3.4.3 Driving Simulator

A National Advanced Driving Simulator MiniSim medium-fidelity driving simulator (Figure 3.3) was used for all driving scenarios. The simulator consisted of three Panasonic TH-42PH2014 42" plasma displays, each with a 1280x800 pixel resolution, which allows 130-degree field of view in front of the seated participant. The center monitor is 28 inches from the center of the steering wheel and the left and right monitors are 37 inches from the center of the steering wheel. The MiniSim 70 also includes 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. 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 records lane position and vehicle speed.

As seen in the Figure 3.3, the gesture control system is positioned to the right of the driver sitting in the driving simulator. The center of the monitor is 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 3.3 driving simulator setup, visual display monitor with webcam, and LEAP Motion.

3.4.4 Eye Tracking

Eye glance behaviors were recorded by a webcam placed on top of the visual display monitor (Figure 3.3). The eye glances were later coded by a researcher and placed into three categories based on the estimated length of the glance duration: short (< 1s), medium (1-2s), and long (> 2s). I 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).

3.4.5 Workload

The NASA -TLX (Hart, 1988) is a widely used subjective workload measure that is comprised of 6 subscales: mental, physical, temporal, effort, performance, and frustration. Each of these subscales is rated on a 20-point scale (1 is low, 20 is high). In Experiment 1 and 2, I removed the temporal demand subscale because the presentation of cues to complete secondary tasks occurred at a fixed rate, and therefore, the results might not reveal insights that could lead to actionable design recommendations. After completing a task, participants rated their perceived workload on each of the six subscales, and then made 15 comparative judgements between pairs of subscales about which was a bigger contributor to their workload, (e.g., more mental or physical?). The Raw TLX (Hart, 2006), is a streamlined version of the NASA-TLX, that excludes the weighting questions. This is a more efficient way to administer the test because it is much shorter. For Experiments 1 and 2, the experimenter administered the Raw TLX, without the temporal subscale. See below for the meanings of each of the subscales

Mental demand – How much mental and perceptual activity was required.

Physical demand – How much physical activity was required.

Temporal demand – How much time pressure the participant felt as a result of the rate/pace of tasks.

Effort – How hard did the participant feel they had to work.

Performance – How successful did the participant feel they were.

Frustration – How insecure, discouraged, irritated, stressed, or annoyed was the participant during the task.

The NASA-TLX is a subjective rating considering all tasks being performed. So, a participant in this study will rate his/her experience with the combination of driving and the gesture controls. The NASA-TLX is not sensitive to differences in perceived workload between the tasks being performed, i.e. driving versus secondary task. This is normal practice for implementation of the NASA-TLX, but it does limit the extent to which one can make statements about which tasks were contributing most to the workload.

3.4.6 Experimental Design

The study was a within-subjects repeated measures factorial design. Each participant completed all four conditions in one session. Each session took about one hour to complete.

- 2x2 grid, Gestures, with Auditory feedback (2x2GA)
- 2x2 grid, Gestures, no auditory feedback (2x2G)
- 4x4 grid, Gestures, with Auditory feedback (4x4GA)
- 4x4 grid, Gestures, no auditory feedback (4x4G)

3.4.7 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).

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 – NASA-TLX (Hart & Staveland, 1988) provides a standardized measure of workload, including measures of physical, cognitive, and temporal demand, as well as perceptions of effort, performance, and frustration.

3.4.8 Procedure

Training

Participants were first trained to use the gesture control systems for five minutes. Participants then practiced driving in the simulator for several minutes to become acclimated and even practiced using the system and driving simultaneously. The participants were given no instructions about how they should balance the demands of the primary and secondary task. Training was done to mitigate as much as possible the learning effects associated with using air gesture control systems.

Prototype systems

The order in which participants used the prototypes was randomized. A total of 32 selection tasks, evenly divided between target options, were completed for each prototype system, taking approximately five minutes to complete. Auditory cues instruct participants which target to select (e.g., “Select option B”). The order of the auditory cues was randomly determined by the Pure Data patch.

Questionnaires

After completing all of the selection tasks, notes were taken about participants’ first impressions. Next, participants were asked several questions about their workload (Hart, 1988) including: mental demand, physical demand, performance, effort, and frustration from the NASA-TLX workload assessment. This process was repeated for all four prototypes.

3.4.9 Statistical Analysis

Repeated measures ANOVAs (2x2 within subjects design) were used to show main effects of Grid Layout and Auditory display factors. Partial eta-squared was also reported as a measure of effect size.

For the driving performance measures (lane departures, standard deviations, and speed) data for two participants was removed from analysis because of partial data loss. For the secondary task measures (time, accuracy) nine participants’ data was removed because of partial data loss. Partial data loss in all cases was due to experimenter error.

3.5 Results

3.5.1 Driving Performance

Lane departures

Repeated measures ANOVA results showed a main effect for the Grid Layout $F(1, 20) = 21.29$, $p < .001$, $\eta_p^2 = .516$ (Figure 3.4). There was also a main effect for Auditory Displays, $F(1,20) = 5.02$, $p = 0.037$, $\eta_p^2 = 0.201$, but no statistical interactions, $F(1,20) = 0.232$, $p = 0.636$, $\eta_p^2 = 0.011$. Paired samples t-tests showed significantly more time spent out of the correct lane for the 4x4GA compared to the 2x2GA, $t(20) = -3.36$, $p = 0.003$, and 2x2G, $t(20) = -2.95$, $p = 0.008$. The 4x4G also led to statistically greater time out-of-lane compared to the 2x2GA, $t(20) = -4.38$, $p < 0.001$, and 2x2G, $t(20) = -3.34$, $p = 0.003$.

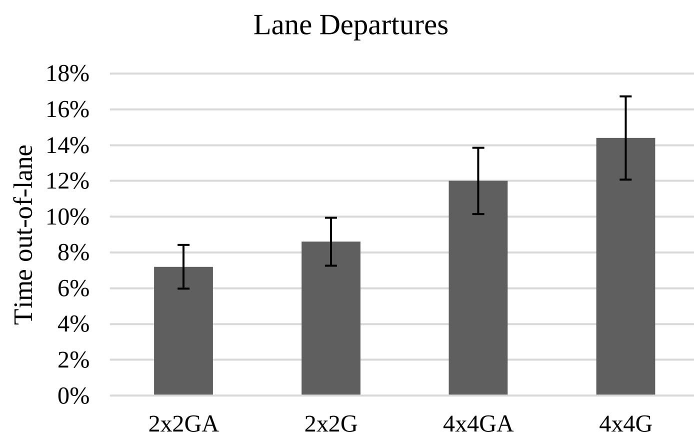

Figure 3.4 percent of time driven outside correct lane. Error bars denote standard errors.

Standard deviation of lane position showed a similar pattern or results (Table 3.3. The Grid Layout menu showed a main effect on standard deviations of lane position, $F(1,20) = 21.052$, $p < 0.001$, $\eta_p^2 = 0.513$, but the Auditory Display did not show a main effect, $F(1,20) = 1.505$, $p = 0.234$, $\eta_p^2 = 0.070$. There was also no statistically significant interaction, $F(1,20) = 0.382$, $p = 0.544$, $\eta_p^2 = 0.019$.

Table 3.2 descriptive statistics for standard deviation of lane position.

	2x2 GA	**2x2 G**	**4x4GA**	**4x4G**
Mean (m)	1.72	1.76	2.02	2.15
SD (m)	0.41	0.35	0.57	0.62

Speed

Participants were instructed to drive between 30-40 mph over the duration of the experiment. The average speed while using each of the four prototypes was between 31-34 mph (Table 3.4). ANOVA results showed a main effect for the Grid Layout (2x2 vs. 4x4) on average speed, $F(1, 20) = 18.7$, $p < .001$, $\eta_p^2 = .483$. Auditory Display did not have a main effect, $F(1,20) = 0.00$, $p = 0.989$, $\eta_p^2 < 0.001$, and interactions did not have a significant effect, $F(1,20) = 0.277$, $p = 0.604$, $\eta_p^2 = 0.014$.

ANOVA of standard deviations in speed also did not show any differences for the Grid Layout, $F(1,20) = 2.158$, $p = 0.157$, $\eta_p^2 = 0.097$, Auditory Display, $F(1,20) = 0.668$, $p = 0.423$, $\eta_p^2 = 0.032$, or statistical interactions, $F(1,20) = 0.008$, $p = 0.930$, $\eta_p^2 < 0.001$.

Table 3.3 means and standard deviations of driver speed

	2x2 GA	**2x2 G**	**4x4GA**	**4x4G**
Mean (mph)	33.5	33.3	31.8	31.9
SD (mph)	3.5	3.3	3.7	3.6

3.5.2 Eye Glances

Short glances (<1s)

ANOVA results showed main effects for both the Grid Layout (2x2 vs. 4x4), $F(1, 20) = 93.9$, $p < .001$, $\eta_p^2 = .824$, and the Auditory Display (present vs. absent), $F(1, 20) = 22.2$, $p < .001$, $\eta_p^2 = .527$ (Figure 3.5). ANOVA results showed no significant interaction between Grid Layout and Auditory Display factors, $F(1,20) = 0.156$, $p = 0.697$, $\eta_p^2 = 0.008$.

Medium glances (1-2s)

ANOVA results showed main effects for both the Grid Layout (2x2 vs. 4x4), $F(1, 20) = 79.7$, $p < .001$, $\eta_p^2 = .799$, and the Auditory Display (present vs. absent), $F(1, 20) = 42.3$, $p < .001$, $\eta_p^2 = .679$ (Figure 3.5). ANOVA also showed an interaction effect, $F(1, 20) = 21.7$, $p < .001$, $\eta_p^2 = .521$. As can be seen in Figure 10, the Auditory Display reduced the frequency of medium glances more for the 4x4 grid than the 2x2 grids.

Long glances (>2s)

ANOVA results showed main effects for both the grid layout (2x2 vs. 4x4), $F(1, 20) = 14.3$, $p = .001$, $\eta_p^2 = .417$, and the Auditory Display (present vs. absent), $F(1, 20) = 9.04$, $p = .007$, $\eta_p^2 = .311$ (Figure 3.5). ANOVA results showed a significant interaction effect $F(1, 20) = 9.04$, $p = .007$, $\eta_p^2 = .311$. Again, the interaction is a result of the auditory impact reducing the number long glances more for the 4x4 grids than the 2x2 grids.

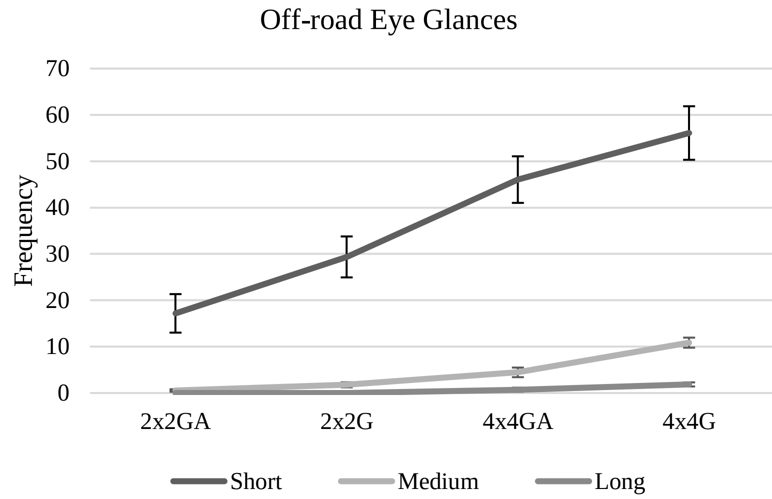

Figure 3.5 average numbers of off-road glances across conditions. Error bars denote standard errors.

3.5.3 Secondary Task Performance

Time

ANOVA results showed a main effect for the Grid Layout, $F(1, 13) = 57.1$, $p < .001$, $\eta_p^2 = .814$ (Table 4). There was no main effect for Auditory Display, $F(1,13) = 0.334$, $p = 0.573$, $\eta_p^2 = 0.025$, or any significant interaction, $F(1,13) = 0.236$, $p = 0.635$, $\eta_p^2 = 0.018$.

Accuracy

ANOVA results showed a main effect for the Grid Layout, $F(1, 13) = 47.8$, $p < .001$, $\eta_p^2 = .786$ (Table 3.5). There was no main effect for Auditory Display, $F(1,13) = 0.013$, $p = 0.911$, $\eta_p^2 = 0.001$, or statistically significant interaction, $F(1,13) = 0.046$, $p = 0.833$, $\eta_p^2 = 0.004$.

Table 3.4 mean selection time and accuracy across condition

	2x2GA	**2x2G**	**4x4GA**	**4x4G**

Time (ms)	2779	2655	4021	3826
Accuracy (%)	91%	89%	63%	67%

Target position

I measured average time to make selections as well as average accuracy for different target positions in 2x2 and 4x4 grids (4x4 results in Figure 3.6). The numbers in Figure 3.6 represent the average of the grids with and without auditory feedback because Auditory Display did not have a significant impact on secondary task performance for 2x2 or 4x4 grids. Note the superior performance of targets in the upper left corner of the 2x2 and 4x4 grids. Also note the relatively poor performance of targets in the lower left corner in the 4x4 grids. The patterns seen in Figure 3.6 were unexpected, but some potential explanations will follow in the discussion section.

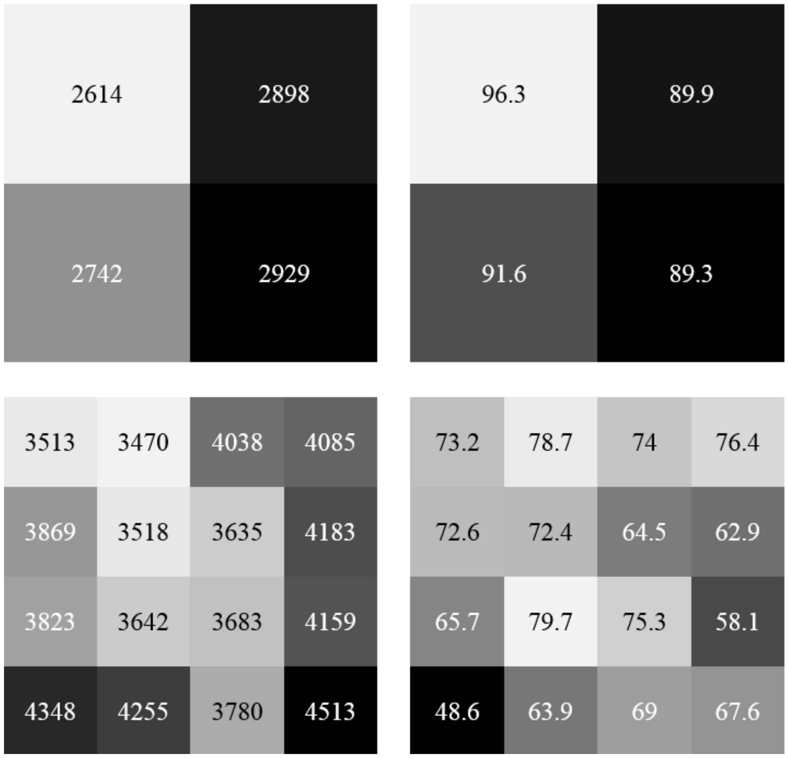

Figure 3.6 (Left) average selection times (ms) for each target position in 2x2 and 4x4 grids. (Right) average accuracy rates (% correct) for each target position in 2x2 and 4x4 grids. Lighter colors indicate faster selection times and higher accuracy rates.

3.5.4 Workload

Physical demand

ANOVA results showed main effects for the Grid Layout (2x2 vs. 4x4), $F(1, 22) = 20.8$, $p < .001$, $\eta_p^2 = .486$ (Figure 3.7). There were no significant main effects for Auditory Display, $F(1,22) = 1.294$, $p = 0.267$, $\eta_p^2 = 0.056$ or interaction effects, $F(1,22) = 1.715$, $p = 0.204$, $\eta_p^2 = 0.072$.

Frustration

ANOVA results showed main effects for the Grid Layout, $F(1, 22) = 37.91$, $p < .001$, $\eta_p^2 = 0.633$, and Auditory Display, F(1,22) = 4.342, p = 0.049, $\eta_p^2 = 0.165$ (Figure 3.7). There were no significant interaction effects, $F(1,22) = 4.021$, $p = 0.057$, $\eta_p^2 = 0.155$.

Mental demand

ANOVA results showed main effects for both the Grid Layout, $F(1, 22) = 91.7$, $p < .001$, $\eta_p^2 = .806$, and the Auditory Feedback, $F(1, 22) = 7.86$, $p < .010$, $\eta_p^2 = .263$ (Figure 3.7). ANOVA also showed an interaction effect, $F(1, 22) = 6.13$, $p < .022$, $\eta_p^2 = .218$. In this case, the auditory feedback reduced the mental demand more for the 2x2 grid than for the 4x4 grid.

Effort

ANOVA results showed main effects for both the grid layout, $F(1, 22) = 24.9$, $p < .001$, $\eta_p^2 = .531$, but not for the Auditory Display, $F(1, 22) = 3.95$, $p = .059$, $\eta_p^2 = .152$ (Figure 3.7). ANOVA also showed an interaction effect, $F(1, 22) = 8.23$, $p = .009$, $\eta_p^2 = .272$. Again, note that the auditory feedback reduces participant effort more for the 2x2 grids than for the 4x4 grids.

Performance

ANOVA results showed main effects for the Grid Layout, $F(1, 22) = 77.4$, $p < .001$, $\eta_p^2 = .778$ (Figure 3.7), but no main effect for Auditory Display $F(1,22) = 0.819$, $p = 0.375$, $\eta_p^2 = 0.036$. There were no significant interaction effects, $F(1,22) = 0.040$, $p = 0.844$, $\eta_p^2 = 0.002$.

3.6 Discussion

Hypothesis 1 was fully supported by the experimental results. The 2x2 grids (fewer,

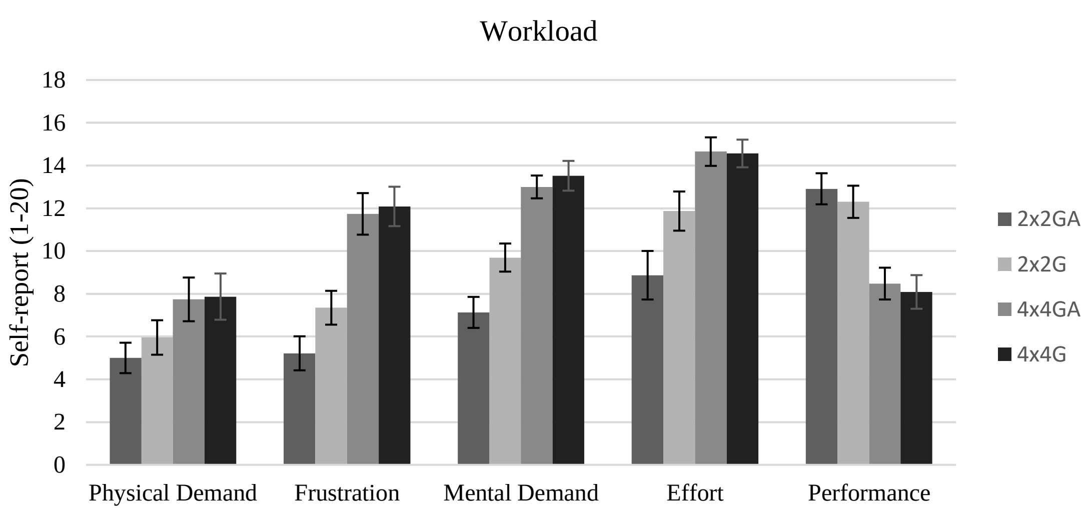

Figure 3.7 NASA TLX subscale results for each prototype. Error bars denote standard errors.

larger targets) resulted in reduced standard deviation of lane position, fewer glances away from the road, and lower workload when compared to the 4x4 grids. Hypothesis 2 was partially confirmed. Auditory feedback reduced the frequency of off-road glances (short, medium, and long) and decreased driver workload. However, auditory feedback did not reduce lane departures or improve secondary task performance.

For several measures including lane departures, secondary task performance, and driver workload, the grid layout had larger effects than the auditory feedback. Results from the driving speed suggest that participants were driving slower while using 4x4 grids to compensate for the difficulty of the secondary task, as has been observed in previous research (e.g., Alm & Nilsson, 1995; Drews, Yazdani, Godfrey & Cooper, 2009). This suggests that the overall task difficulty was greater for the 4x4 grids, regardless of presence of an auditory display. The number of lane departures was also higher for the 4x4 grids compared to 2x2 grids.

Although the grid layout appears to have a larger effect across nearly all measures, auditory displays also impacted many measures. Auditory displays dramatically reduced the number of off-road glances for both grid layouts and driver workload. There was a statistical interaction showing that auditory displays had a larger effect in reducing the visual demands for the 4x4 grids than the 2x2 grids. This interaction can be explained by the relatively higher visual demand required to complete target selections with the 4x4 grids. While 2x2 grids had only 4 menu items, whose positions could be easily memorized and located, the 4x4 grids had 16 menu items, each of which could not be easily memorized. This means that target selections using the 4x4 grids required more searching because the interface contains more information, making the secondary task is relatively more difficult, compared to the 2x2 grid selections. It also means that the bandwidth requirements to make quick selections was also higher, suggesting that drivers would benefit from relying on the visual display. However, because priority was placed on the driving task, drivers were often forced to rely more on the auditory display to avoid visual conflict. This observation supports the idea that auditory displays can be of relatively greater benefit, i.e., reduce the number of off-road glances, for more visually demanding secondary tasks, such as those observed when participants completed target selections using 4x4 grid prototypes. Interestingly, the addition of auditory displays reduced the visual demands required to complete the secondary task but did not improve driving performance. This suggests that participants' ability to successfully balance primary and secondary tasks was not solely influenced by the competition for visual resources. Even though the results did not reach statistical significance, it is possible to see in Figure 9 that participants' time-out-of-lane showed that adding auditory displays were associated with reduced time-out-of-lane for both 2x2 and 4x4 grids. Upon closer examination, it is also possible to see that the impact of adding auditory displays is about twice as big (although not statistically significant) for the 4x4 condition as the 2x2 condition. This pattern is consistent with the interaction observed for the visual displays. It remains a possible explanation that visual demand mediates, in part, the relationship between auditory displays and driving performance. As Wickens' Multiple Resource Theory suggests, secondary task difficulty can also influence a performer's ability to multi-task (Wickens, 2002). The difference in task difficulty between selecting the large targets in the 2x2 grid and small targets in 4x4 grids may have a larger impact on multitasking performance than the reduction in competition for visual resources. Another possibility is that the driving environment did not require sufficient visual resources to be sensitive to the difference in availability of vision between systems with auditory feedback and those without. This could be possible because the driving scenario had no other vehicles, and traffic signs or signals, and was a small closed loop, which was repeated continuously. Despite the lack of statistically significant improvements in driving performance, the reduction in eyes-off-road time can facilitate improve situation awareness and increase drivers' ability to respond to hazardous situations on the road.

Participants' comments during the experiment revealed that the auditory feedback was helpful for the 2x2 grids, assuring them of the systems status, but for the 4x4 grids the auditory feedback was more disruptive than helpful. Due to the large number of targets the auditory feedback became noisy and difficult to understand, rather than a signal of the system status. These comments are consistent with the trends, and statistical interactions observed in the workload measures. In other words, participants found the auditory feedback reduced their mental and physical demand, effort, and frustration, and improved their performance using the 2x2 grids, but the same auditory feedback led to little or no improvement in the 4x4 grids. These results are interesting because they apparently contradict the observations that adding auditory displays further reduce off-road glances and time-out-of-lane, for 4x4 grids compared to 2x2 grids. It appears that, in the case of workload, the impact of grid layout is a greater factor than display modality factor. In other words, perceptions of workload across most of the subscales were more dramatically impacted by the size and/or the number of targets than the presence of auditory displays, despite the greater reduction of visual demand associated with visual displays.

When analyzing the data, I explored the effect of target position on secondary task performance. My initial assumption was that closer targets, with lower *indices of difficulty*, IDs, should result in faster selections and higher accuracy, and further targets with higher IDs should have slower selection times and lower accuracy rates. Generally this held true, but I found that there was an arc across the 4x4 grid – which was not noticeable in the low granularity of the 2x2 grid – along which selections were faster and accuracy was higher. Targets which were among the closest, at the bottom left corner of the 4x4 grid, resulted in low accuracy and slower selections. My interpretation of this result, since the effect appears for both speed and accuracy, is that there is a bubble in space which the operator can reach, and within that bubble there are some places that are harder to complete otherwise equivalent target acquisition tasks. It is possible that the sensor position was such that participants found it more difficult to select targets that were especially close to their bodies, as well as targets that were especially far away. This result highlights a need for further research investigating in-air target acquisition performance within different areas of the reach envelope of users.

To help other researchers who may be interested in what target sizes to use for eyes-free interaction, I estimated the target sizes and calculated the index of difficulty (ID) of the 2x2 and 4x4 grids. In this case, the 2x2 grid targets had a Fitts's ID from a range of 1.77 to 2.13, and the 4x4 grid targets had an ID ranging 2.43 to 3.07.

Chapter 4

4 Experiment 2 – Comparing Air Gesture Controls to Touchscreens

4.1 Introduction

I wanted to compare the best gesture control prototypes to an equivalent touchscreen system in order to determine what, if any, benefits gesture controls provide over touchscreens. I evaluated 2x2 grids with and without auditory feedback, and compared them to 2x2 grids on a touchscreen. The procedure was the same as Experiment 1.

4.2 Hypotheses

H1: Touchscreen use will be more visually demanding than the gesture controls with auditory feedback and will result in higher frequencies of off-road glances, especially for glances less than one second. Gesture controls without auditory feedback will result in the most off-road glances.

H2: Touchscreen use will degrade driving performance more than gesture controls with auditory feedback, but less than gesture controls without auditory feedback. I anticipate both time out-of-lane and variance in car following distance will be greater in the touchscreen conditions than the gesture controls with auditory feedback, but all conditions will be better than the gesture controls without auditory feedback.

4.3 Methods

4.3.1 Experimental Design

The study was a within-subjects repeated measures factorial design. Each participant completed all four conditions in one session. Each session took about one hour to complete.

- 2x2 grid, Gesture, Auditory feedback (2x2GA)
- 2x2 grid, Gesture, no auditory feedback (2x2G)
- 2x2 grid, Touchscreen, Auditory feedback (2x2TA)
- 2x2 grid, Touchscreen, no auditory feedback (2x2T)

4.3.2 Participants

A total of 24 participants, 13 males and 11 females, were recruited from the undergraduate psychology student pool at Michigan Technological University (Table 4.1). None of the participants participated in the previous experiment.

Table 4.1 participant demographics

	Age	Experience (yrs)	Miles/yr
Mean	20.21	3.98	7538
SD	1.86	1.49	6637

4.3.3 Apparatus

A small touchscreen-capable laptop (10.1 inch screen, 1280x800 resolution) was used as the touchscreen. It was positioned directly in front of the monitor pictured in Figure 3.3, but placed on a stand to position the touchscreen at the midpoint of the computer monitor so that touchscreen and air gesture prototypes were displayed from the same position. The touchscreen computer was not used for all conditions because it was not capable of running the LEAP Motion software without significant lag.

4.3.4 Methodological Differences

After completion of Experiment 1, I decided to add a lead vehicle to the driving scenario. Participants were instructed to follow the lead vehicle at a constant safe distance. The speed of the vehicle varied over time. The lead vehicle speed changed every 10 seconds. Its speed was determined by sampling from a normal distribution with a mean of 33 miles per hour and standard deviation of 7 miles per hour. This methodological change normalizes the speed of the drivers, making the task difficulty more consistent because in Experiment 1 some participants chose to drive slower, possibly as a compensatory action to reduce their workload. Adding a lead vehicle also requires participants to track the distance to a lead vehicle in addition to lane keeping, making the task overall more difficult and more representative of real-world driving. Distance from the driver to the lead vehicle can be used as a measure of task difficulty because drivers tend to follow at greater distances to reduce their workload (Strayer & Drew, 2004) and variance in following distance can be interpreted as a measure of the driver's ability to attend and react to relevant changes in the driving environment.

The visual displays of the grids were made smaller in Experiment 2 (reduced from 15x15 inches to 4x4 inches). I reduced the size in order to make the experimental conditions more ecologically representative than in Experiment 1, where the visual display was very large. In addition, the angle of the monitor was changed to be parallel to the wheel, and the simple target labels, "A, B, C, D" were changed to be "audio, navigation, phone, settings", again, in an effort to be more representative of real-world in-vehicle control setups.

Training followed the same method as in Experiment 1, but training time was split between touchscreens and air gesture control systems, spending the majority of time working on air gesture controls, roughly split 4 minutes to 1 minute for air gestures to touchscreens. Training was done to mitigate learning effects as much as possible, which is especially important when making comparisons to touchscreens, a much more familiar input method.

4.4 Results

4.4.1 Driving Performance

Lane departures

Drivers using the gesture control systems drove out of their lane more than drivers using touchscreens (Table 4.2). ANOVA results show that the Input Method (Gesture vs. Touch) had a main effect on lane departures $F(1, 21) = 10.9$, $p = .003$, $\eta_p^2 = .342$. Auditory Display did not show a main effect on lane departures, $F(1,21) = 0.016$, $p = 0.901$, $\eta_p^2 = 0.001$, and there were no significant interactions, $F(1,21) = 0.939$, $p = 0.344$, $\eta_p^2 = 0.043$.

Following distance

Drivers tended to follow lead vehicles more closely when using the touchscreens than when using gesture controls (Table 4.2). This suggests that participants may feel lower workload when using the touchscreens when compared to the gesture controls. However, ANOVA results showed no statistically significant main effects for Input Method,

F(1,23) = 3.752, p = 0.065, η_p^2 = 0.140, or Auditory Display, F(1,23) = 0.158, p = 0.694, η_p^2 = 0.007, or statistical interactions, F(1,23) = 0.010, p = 0.923, $\eta_p^2 < 0.001$.

Following distance also varied more when drivers were using the gesture controls than the touchscreen controls (Table 4.2). However, Input Method did not reach statistical significance according to ANOVA results, F(1,21) = 1.170, p = 0.292, η_p^2 = 0.053. There was also no main effect for Auditory Display, F(1,21) = 3.527, p = 0.074, η_p^2 = 0.144, or statistical interaction, F(1,21) = 0.430, p = 0.519, η_p^2 = 0.020.

Table 4.2 means of driving performance measures

	2x2GA	**2x2G**	**2x2TA**	**2x2T**
Lane Dep. (%)	3.0%	2.8%	1.2%	1.5%
Dist. (ft)	137	141	128	132
SD of Dist. (ft)	54.5	58.1	51.2	54.2

4.4.2 Eye Glance Behavior

Short glances (<1s)

The 2x2GA system resulted in fewer off-road glances than any other systems (Figure 4.1). Meanwhile, the 2x2G was the most visually demanding. ANOVA results suggest Auditory Display influenced eye glances F(1, 22) = 45.7, $p < .001$, η_p^2 = .675, but Input Method did not F(1, 22) = .622, p = .439, η_p^2 = .028. There was a significant interaction between the two factors, F(1, 22) = 54.2 $p < .001$, η_p^2 = .711. This interaction can be seen in Figure 13 below, which shows the Auditory Display has a large influence on the gesture controls but no influence on the touchscreen controls.

Medium glances (1-2s)

Touchscreens generally resulted in fewer medium glances than gesture controls (Figure 4.1). Overall, there were very few medium glances for any of the systems. ANOVA results showed the Input Method impacted the number of medium glances $F(1, 22) = 24.4$, $p < .001$, $\eta_p^2 = .526$. Auditory Display also showed a main effect, $F(1, 22) = 23.2$, $p < .001$, $\eta_p^2 = .513$, and there was a significant interaction between those two factors, showing that Auditory Display reduced visual demands for the gesture controlled systems, but was not important for touchscreens $F(1, 22) = 23.4$, $p < .001$, $\eta_p^2 = .516$. Despite the small number of glances between 1 and 2 seconds, a consistent pattern emerged showing the 2x2G system as more visually demanding than all other prototypes.

Long glances (>2s)

Only the gesture control system without auditory feedback resulted in any long glances. However, there were no main effects of Input Method, $F(1,22) = 2.095$, $p = 0.162$, $\eta_p^2 = 0.087$, Auditory Displays, $F(1,22) = 2.095$, $p = 0.162$, $\eta_p^2 = 0.087$, or statistical interactions, $F(1,22) = 2.095$, $p = 0.162$, $\eta_p^2 = 0.087$.

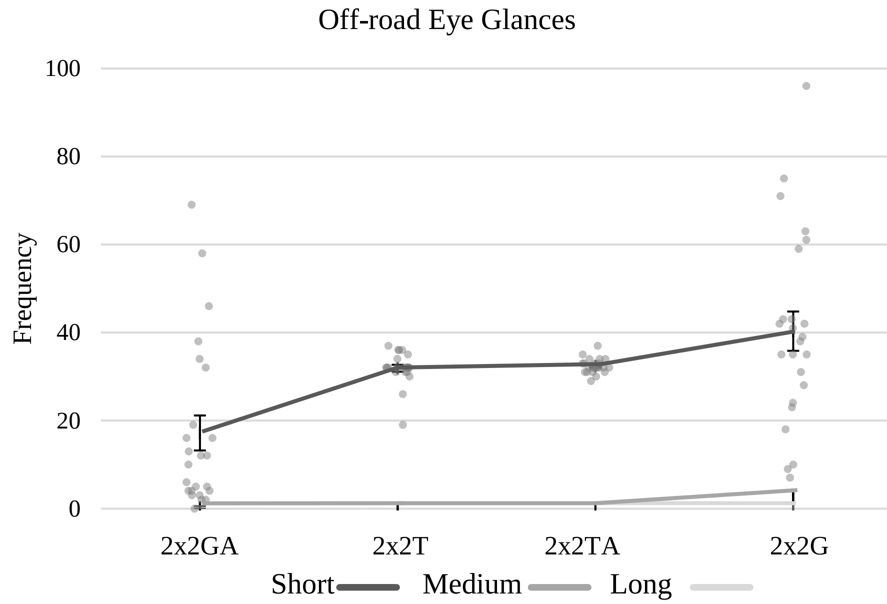

Figure 4.1 average number of off-road glances across conditions. Points jittered represent short glance counts. Error bars denote standard errors.

4.4.3 Secondary Task Performance

Time

Average time to make a selection using the gesture controls was significantly slower than selections made using the touchscreen (Table 4.3). ANOVA results show that Input Method impacted the selection times $F(1, 23) = 186$, $p < .001$, $\eta_p^2 = .894$, but Auditory Display did not, $F(1,23) = .244$, $p = .626$, $\eta_p^2 = .011$. There was no significant interaction between factors, $F(1,23) = 2.792$, $p = 0.109$, $\eta_p^2 = 0.113$. Selection times were not significantly different between the two gesture control systems or the two touchscreen control systems.

Accuracy

Average selection accuracies were higher for touchscreens than gesture controls. Touchscreens reached nearly perfect levels of accuracy, with gesture controls performing

5-6 percentage points worse (Table 4.3). ANOVA results showed that Input Method had significant effects on selection accuracy, $F(1, 23) = 35.9$, $p < .001$, $\eta_p^2 = .609$, as well as Auditory Display, $F(1, 23) = 9.66$, $p = .005$, $\eta_p^2 < .296$. There were no statistically significant interactions, $F(1,23) = 0.848$, $p = 0.367$, $\eta_p^2 = 0.036$.

Table 4.3 Mean selection times and accuracy rates across all conditions

	2x2GA	**2x2G**	**2x2TA**	**2x2T**
Time (ms)	2832	2758	1967	1996
Accuracy (%)	91%	94%	97%	99%

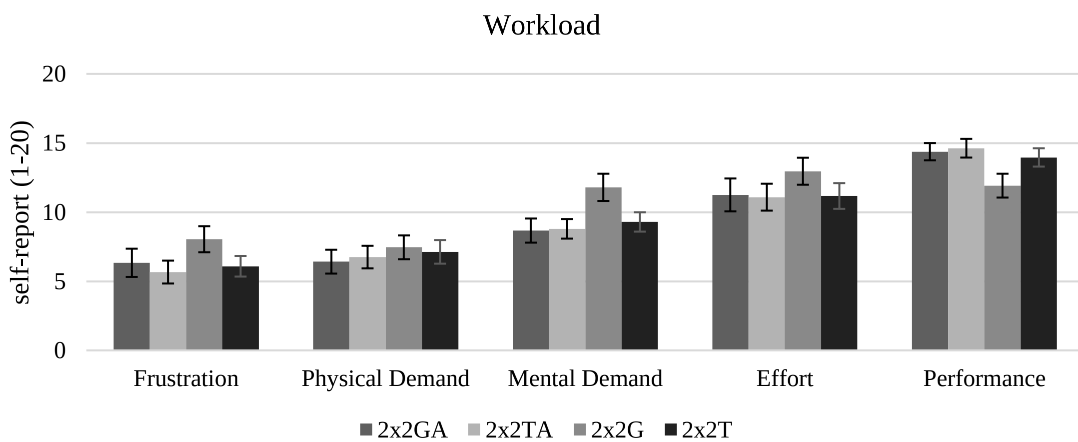

Figure 4.2 NASA TLX subscale results for each prototype design. Error bars denote standard errors.

4.4.4 Workload

Frustration

ANOVA results showed that neither Input Method, $F(1,23) = 2.896$, $p = 0.102$, $\eta_p^2 = 0.112$, nor Auditory Display, $F(1,23) = 2.611$, $p = 0.12$, $\eta_p^2 = 0.102$ significantly impacted perceptions of frustration. There were also no significant interactions, $F(1,23) = 1.939$, $p = 0.177$, $\eta_p^2 = 0.078$ (Figure 4.2)

Physical demand

ANOVA results showed that neither Input Method, $F(1,23) = 0.00$, $p = 1.00$, $\eta_p^2 < 0.001$ nor Auditory Display, $F(1,23) = 2.58$, $p = 0.122$, $\eta_p^2 = 0.101$ significantly impacted perceptions of physical workload. There were also no significant interactions, $F(1,23) = 1.282$, $p = 0.269$, $\eta_p^2 = 0.053$ (Figure 4.2).

Mental demand

ANOVA results showed that Input Method $F(1, 23) = 18.2$, $p < .001$, $\eta_p^2 = .441$, impacted mental demand (Figure 4.2). There was also a significant interaction between Input Method and Auditory Display, $F(1, 23) = 10.2$, $p = .004$, $\eta_p^2 = .307$. The interaction likely represents that there is increase in mental demand associated with removing auditory displays from the gesture control system, but the same increase is not seen when removing the auditory display from the touchscreens.

Effort

ANOVA results showed both Input Method, $F(1,23) = 1.65$, $p = 0.212$, $\eta_p^2 = 0.067$, and Auditory Display, $F(1,23) = 3.602$, $p = 0.070$, $\eta_p^2 = 0.135$ did not significantly impact perceptions of effort. There were also no statistically significant interactions, $F(1,23) = 3.002$, $p < 0.097$, $\eta_p^2 = 0.115$ (Figure 4.2)

Performance

Performance ratings were lower for the 2x2G system than any other systems (Figure 4.2). ANOVA results showed that Input Method, $F(1, 23) = 7.52$, $p = .012$, $\eta_p^2 = .246$, impacted performance ratings. There was also a significant interaction between the Input Method and Auditory Display factors, $F(1, 23) = 4.90$, $p = .037$, $\eta_p^2 = .175$. This interaction is likely reflecting the drop in performance associated with removing the auditory feedback for the gesture control prototypes, an effect which was not observed for the touchscreen systems.

4.5 Discussion

The goal of Experiment 2 was to take the best performing prototypes from Experiment 1, the 2x2 grids with and without audio feedback, and compare those prototypes to equivalent touchscreen system in a within-subjects experimental design. Below I review the results in light of the hypotheses proposed before the experiment was conducted.

Hypothesis 1 said that there would be more off-road glances for touchscreens than air gesture controls with auditory feedback, but air gesture controls without auditory feedback would result in the most off-road eye glances. Hypothesis 1 was supported. The data showed that gesture controls with auditory feedback resulted in fewer off-road glances, followed by the two touchscreen systems, with gesture controls without auditory feedback requiring the most off-road glances. Notably, all systems performed within NHTSA guidelines (National Highway Traffic Safety Administration, 2012).

Hypothesis 2 stated that air gesture controls with auditory feedback would lead to more time-in-lane, and a shorter following distance to the lead vehicle compared to touchscreen conditions, but that air gesture controls without auditory feedback would result in the worse driving performance than all other systems. Hypothesis 2 was not supported. While variance in following distance was equivalent between touchscreens and air gesture controls, gesture controls led to less time-in-lane. The literature that supported hypothesis 2 suggested that by reducing the visual demand of the secondary task, driving performance could be improved. Since it was still observed in results from Experiment 2 that visual demand was decreased, there must be alternative explanations for the lack of driving performance improvement. One possible explanation is that target selections with the gesture control system were more difficult, requiring more mental or physical resources to complete. However, subjective workload results showed no significant differences between 2x2GA, 2x2TA, and 2x2T prototypes, meaning that participants did not perceive greater workload for air gesture controls with auditory displays. It remains possible that the participants' perceptions of workload are not accurate reflections of their real workload. Another possible explanation is that target selections took longer for the gesture controls. The reason for this could be due to limited practice time with a novel system, or the relatively limited information capacity of the combination of auditory and proprioceptive modalities. In any case, the target selections took longer when using the 2x2GA prototype meaning that participants are dividing their attentional resources between the primary and secondary tasks for a longer time compared to the touchscreen systems, which while taking more focal visual attention on average also took less time, meaning that during a greater percentage of their drive duration drivers using the 2x2TA and 2x2T systems were able to dedicate all of their attention and resources to the primary task. Drivers using the 2x2GA system would necessarily be driving with one hand on a curvy road, and dedicating mental resources to searching through the gesture menu for about 40% longer compared to touchscreen use. This explanation implies that driving performance may potentially be degraded, albeit slightly, when driver attention is split between the primary and secondary tasks, despite the improvements in focal visual attention. This explanation undermines the NASA-TLX data which indicated that participants felt no greater mental demand when using the 2x2GA prototype, but does offer an explanation that explains the result and is consistent with Wickens' multiple resource theory.

Why wasn't the addition of the auditory display for other conditions helpful? For the touchscreen interface, the auditory display was only providing feedback, i.e., information *after* completion of the task. It is possible that this feedback could be helpful in guiding future movements as was seen in Hatfield, Wyatt, and Shea (2010). However, that experiment was a reciprocal tapping task, requiring rapid, continuous movement whereas the target acquisitions in Experiment 2 were discrete. In the case of the air gesture control conditions, the audio also did not show improvement. One might expect to see improvements based on previous research showing that adding auditory displays reduced searching times in list navigations (Zhao et al., 2007). However, it is noteworthy that the target acquisition task used in this dissertation does not require a very difficult search subtask. Participants were given time to familiarize themselves with the technology. Additionally, the positions and names of the targets remain fixed throughout the experiment. This means that participants could easily memorize the relative position of the four targets. Therefore, adding auditory information was not helpful in improving accuracy. However, the slower times suggested that participants were using the auditory display to guide their movements. These results are consistent with the statistical interaction observed in Experiment 1 that showed reductions in visual demands associated with the addition of auditory displays were smaller for the 2x2 displays than the 4x4 displays. The 2x2 task is easier, requires less visual demand to search and select intended targets, and therefore cannot benefit as greatly from the addition of auditory displays compared to more difficult target selections.

Variance in following distance was statistically equivalent between air gesture controls and touchscreens. This result suggests that drivers are able to respond to the changes in speed from the lead vehicle just as well when using the gesture controls compared to touchscreens. Even though the percentage of time out of the correct lane was higher when using air gesture controls, the difference was relatively small, increasing from two percent of overall time to three percent of overall time. This is in stark contrast with Experiment 1, in which 4x4 grids led to lane departures covering almost 15% of the drive.

Workload assessments showed that touchscreens, both with and without auditory feedback, were equivalent to the gesture controls with auditory feedback. However, gesture controls without auditory feedback led to higher workload across most measures. Following distance between drivers and the lead vehicle was also statistically equivalent between all conditions, suggesting drivers did not feel a need to compensate to reduce the driving task difficulty (Alm & Nilsson, 1995; Drews, Tazdani, Godfrey, & Cooper, 2009).

A deeper look at the eye glance data shows that there is a much higher variance in the number of glances among the gesture control systems than the touchscreens. Touchscreens, for the most part, required only a single glance to select a target for every participant. However, for the gesture controls, many participants hardly looked away from the road at all, while others looked much more than they did using the touchscreens. It appears there are some individual differences in a desire for visual information that may impact driving performance, secondary task performance, and workload, as well as their willingness to accept auditory-supported gesture control technology. These individual differences in glance behavior could be explained by pre-existing individual differences in our sample group in trust in technology – leading some participants to distrust a new technology – or multitasking prioritization – with some participants placing a relatively higher importance on accuracy in the secondary task. More research is needed to unpack the underlying mechanisms of this relationship.

Selection time and accuracy for the secondary task showed superior performance for touchscreens relative to the gesture controls. This highlights an apparent tradeoff between safety and efficiency, wherein touchscreens appear easier and gesture controls are less visually demanding. The reason that touchscreens are easier is up for debate. One potential explanation is that participants had more familiarity with touchscreen use. Touchscreens are ubiquitous and likely used on a daily basis by most college students, such as the participants in this study. Meanwhile, gesture controls, outside of the use of Nintendo Wii controllers are not used often, and are very unlike the gesture controls presented in the study. The difference in familiarity could explain some of the difference in secondary task accuracy and selection time. The same type of pattern, a safety-secondary task efficiency tradeoff, was observed in May, Gable and Walker (2014). Although their study had a different type of menu structure, they made a direct comparison with a touchscreen system and found that drivers showed safer eye glance behavior but at the cost of slower selection times and greater workload. Their results are largely consistent with the observations from Experiment 2 with the exception of the workload measures. The most likely explanation for the inconsistency in workload measures is that the menu system used in Experiment 2 was simpler, and therefore not sensitive to subjective ratings of workload when compared to the menu used in the May study.

Another important factor to consider is the difference in spatial and temporal resolution between the two sets of equipment. Touchscreens allow for very precise spatial resolution, and because the touchpoints only occur on selection, temporal resolution is not a major factor. However, air gesture control systems present temporal lag issues suggest that participants may not have trusted the system enough to make fast movements. The spatial resolution, while acceptable, was worse than with touchscreens, the sensor may miss the hand position by one or more centimeters, making selection endpoints more variable, and more likely to miss compared to touchscreens.

When participants were asked to rank the four systems the 2x2GA (58%) was ranked first more than the 2x2G (0%), 2x2TA (21%), and 2x2T (21%). The most cited reason for choosing the 2x2GA system was participants' feeling they could attend to the road more easily.

Chapter 5

5 Experiment 3 – Measuring Impact of Control Orientation and Auditory-only Menus

5.1 Introduction

To this point, I have completed research projects comparing 2x2 and 4x4 grids, and comparing in-air gesture controls to touchscreens. However, there are still fundamental questions related to the design of in-vehicle gesture controls. I am interested in understanding the impact of menu orientation. What happens when the mapping is direct, i.e., upwards movements move up on the menu, as opposed to moving forward to move up in the display, analogous to a computer mouse? What is the impact on driving performance and secondary task performance? Another important question is the influence of an auditory-only display on driving performance, secondary task performance, and driver workload. All conditions in the previous two experiments included a visual display along with the auditory display. So far, I have demonstrated that the introduction of auditory displays reduces visual demand of secondary task. However, the eye-glance measures have only considered focal vision. The removal of visual displays will potentially show the impact of peripheral vision on secondary task performance. The removal of visual displays will show the impact of each modality and their interactions.

This experiment was a follow-up to Experiments 1 and 2 in which we found that auditory displays could lead to more eyes-on-road time with limited sacrifice to driving performance. The purpose of this experiment was to learn about the impact of control orientation and to learn the impact of combinations of visual and auditory displays.

5.1.1 Hypotheses

H1: Auditory-only menus will lead to slower and less accurate target selections than visual-only conditions, and visual-auditory displays.

H2: The frequency of off-road glances will be lower in the conditions with the visual-auditory display than the visual-only display, as was observed in Experiments 1 and 2. Auditory-only displays should result in zero, or almost zero off-road eye glances because the air gesture prototypes offer no visual information for the auditory-only prototypes.

H3: The average distance to lead vehicle will be highest for the auditory-only conditions compared to the visual-auditory conditions 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 conditions, visual-auditory conditions, and lowest for the auditory-only conditions.

H4: Vertical menus will lead to higher driver workload, especially for physical demand, compared to the horizontal menu orientation. The vertical menus will also lead to a higher percentage of correct selections when compared to conditions with horizontal orientations.

5.2 Methods

5.2.1 Experimental Design

This was a within-subjects study design. There were a total of six conditions (Table 5.1). 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 5.1 experimental conditions

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

5.2.2 Apparatus

I used a LEAP Motion as our hand-position tracking sensor and I used Pure Data – an open source graphical programming language – to develop our target selection task (Figure 5.1). As the participant moves their hand above the sensor, a cursor matches the position of the person's hand along the x-axis (no y-axis data was recorded) and makes corresponding movements on the screen. All cursor movements are mapped one-to-one to hand movements.

The driving simulator used in this experiment was the same as the previous two experiments. The driving scenario was identical to the one used in Experiment 2.

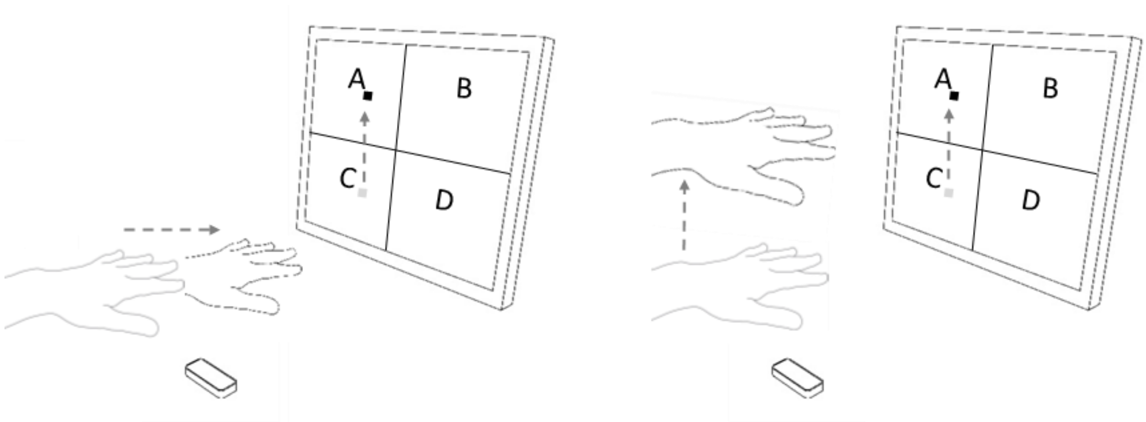

Figure 5.1 illustration of apparatus used showing horizontal (left) and vertical (right) control orientations.

5.2.3 Participants

A total of 24 undergraduate psychology students (21 males, 3 females) were recruited to complete Experiment 3 (Table 5.2). All participants were given course credit as compensation for their participation. Only one person reported having experience using a LEAP Motion before. None of the participants who completed this study had participated in any of the previous experiments. One participant was removed from analysis for the secondary task performance due to data loss.

Table 5.2 participant demographics

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

5.2.4 Procedure

Participants were first 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 is done to mitigate as much as possible the learning effects associated with using a totally novel air gesture control system, an especially important factor when comparing to the much more familiar input method of touchscreens. Participants then practiced driving in the simulator for several minutes to become acclimated and even practiced using the system and driving simultaneously. The participants were given no instructions about how they should balance the demands of the primary and secondary 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 washes 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. Auditory 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, 1988) including: mental demand, physical demand, performance, effort, and frustration from the NASA-TLX workload assessment. This process was repeated for all four prototypes.

5.2.5 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 accuracy and time, one participant's data were removed from analysis because data were missing due to experimenter error.

5.3 Results

5.3.1 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-only displays than Visual-only display, $t(23) = 3.168$, $p = 0.008$, but there were no significant differences between Auditory-only displays and Visual/Auditory, $t(23) = 2.220$, $p = 0.063$, and Visual/Auditory displays and Visual-only displays, $t(23) = -1.828$, $p = 0.074$ (Table 5.5; Figure 5.2).

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-only displays compared to the Visual-only displays, $t(23) = 5.120$, $p < 0.001$, and Visual/Auditory displays, $t(23) = 2.967$, $p = 0.009$. The Visual-only display was not statistically different from Visual/Auditory displays, $t(23) = -2.203$, $p = 0.033$ (Table 5.4).

Table 5.3 means and standard deviations for the standard deviation of lane position

	HA	**HV**	**HVA**	**VA**	**VV**	**VVA**
mean (m)	1.357	1.484	1.401	1.277	1.445	1.396
sd (m)	0.297	0.314	0.266	0.243	0.323	0.253

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 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 display. Meanwhile, the orientation of the control had no impact on lane control.

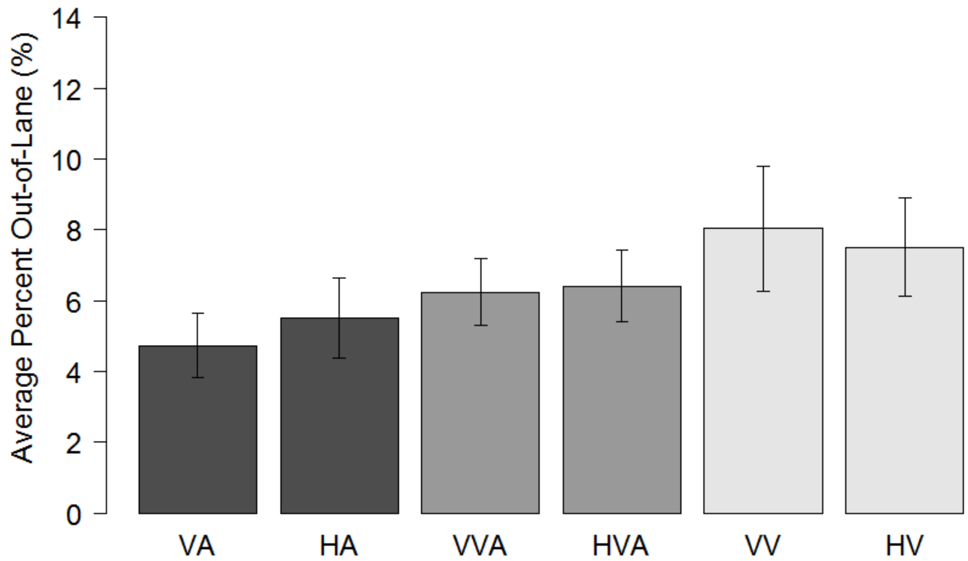

Figure 5.2 Average percentage of time spent out-of-lane.

5.3.2 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 distance for the Visual-only displays compared to the Auditory-only displays, $t(23) = 3.505$, $p = 0.003$. But there were no significant differences between Visual-only and Visual/Auditory displays, $t(23) = -1.692$, $p = 0.195$, or Auditory-only and Auditory/Visual displays, $t(23) = 0.962$, $p = 0.341$ (Table 5.7).

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 the Display factor and Orientation, $F(2,46) = 2.272$, $p = 0.115$, $\eta_p^2 = 0.092$. Paired comparisons showed no significant differences between Visual-only and Visual/Auditory display, $t(23) = -1.372$, $p = 0.530$, or Visual-only and Auditory-only display, $t(23) = -0.150$, $p = 0.881$, or Auditory-only and Visual/Auditory display, $t(23) = 1.360$, $p = 0.530$.

Table 5.4 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

5.3.3 Eye Glances

Short glances (<1 seconds)

ANOVA results showed there was no significant main effect of Orientation on 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-only conditions compared to Visual-only displays, $t(23) = 19.031$, $p < 0.001$, and Visual/Auditory displays, $t(23) = 7.315$, $p < 0.001$. The Visual-only displays led to more off-road eye glances compared to the Visual/Auditory displays, $t(23) = -10.783$, $p < 0.001$ (Figure 5.3).

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 5.3.

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 frequency of 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-only conditions compared to the Visual-only conditions, $t(23) = 6.705$, $p < 0.001$, and Visual/Auditory displays, $t(23) = 4.868$, $p < 0.001$. The Visual-only condition resulted in more off-road eye glances compared to Visual/Auditory conditions, $t(23) = -5.452$, $p < 0.001$ (Figure 5.3).

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-only displays compared to Visual-only displays, $t(23) = 2.808$, $p = 0.022$. But there were no significant differences between Visual-only displays and Visual/Auditory displays, $t(23) = -2.185$, $p = 0.055$, or Auditory-only displays and Visual/Auditory displays, $t(23) = 2.280$, $p = 0.055$.

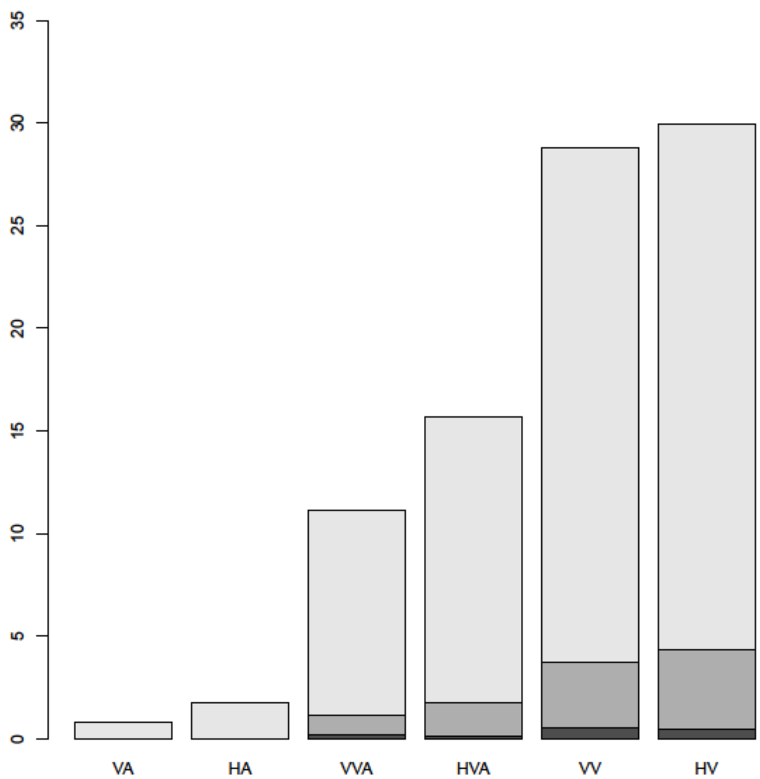

Figure 5.3. Eye glance frequency for short, medium, and long glances.

Table 5.5 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

5.3.4 Time

ANOVA results showed no significant effect of Orientation on selection times, $F(1,22)$ = 0.778, p = 0.387, η_p^2 = 0.034. The Display condition had a significant impact on selection times, $F(2,44)$ = 23.93, $p < 0.001$, η_p^2 = 0.521. There was no statistically significant interaction between Orientation and Display, $F(2,44)$ = 0.097, p = 0.908, η_p^2 = 0.004. Pairwise t-tests showed significant differences between all combinations of displays: Visual/Auditory displays were slower than Visual-only, $t(22)$ = 2.550, p = 0.014, but faster than and Auditory-only displays, $t(22)$ = -5.389, $p < 0.001$. Auditory-only displays were slower than Visual-only displays, $t(22)$ = -7.333, $p < 0.001$ (Table 5.6).

Table 5.6 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

5.3.5 Accuracy

ANOVA results showed no effect of Orientation on task completion accuracy, $F(1,22)$ = 0.875, p = 0.36, η_p^2 = 0.038. The Display factor had no significant effect on task accuracy, $F(2,44)$ = 0.3, p = 0.742, η_p^2 = 0.013. There were no statistically significant interactions between Orientation and Display, $F(2,44)$ = 0.571, p = 0.569, η_p^2 = 0.025. All conditions resulted in mean accuracy rates between 88-92% (Table 5.7).

Table 5.7 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%

5.3.6 Workload

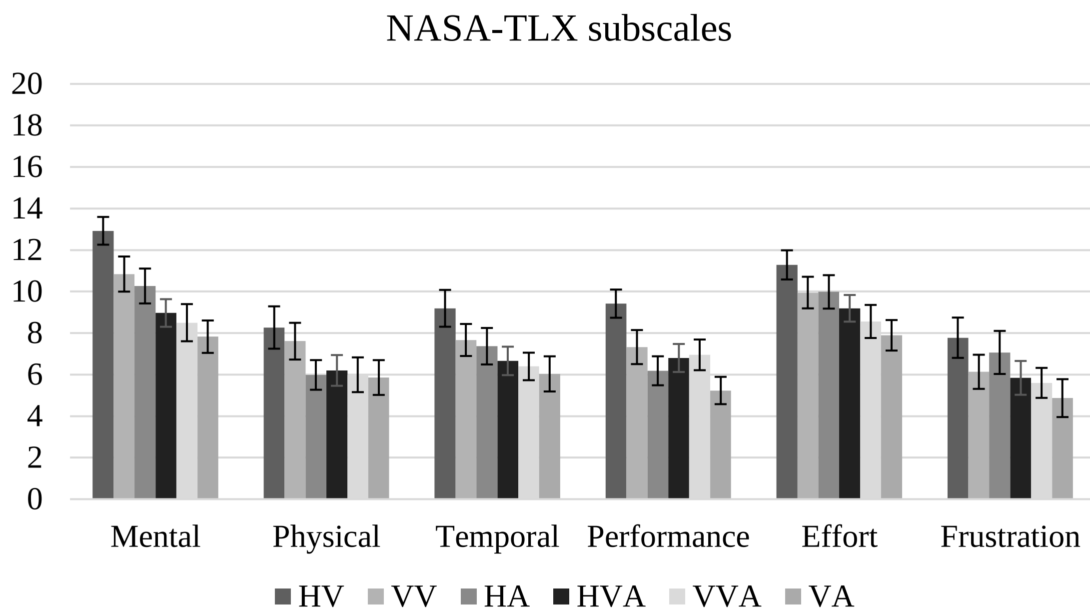

Figure 5.4. NASA-TLX workload subscales.

Mental workload

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 5.4). 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, $F(2,46) = 2.299$, $p = 0.112$, $\eta_p^2 = 0.091$. The Display factor did not have a significant interaction with Orientation, $F(1,23) = 0.457$, $p = 0.506$, $\eta_p^2 = 0.019$. Pairwise t-tests showed the Visual-only conditions led to significantly higher perceived mental workload compared to Visual/Auditory, $t(23) = 5.565$, $p < 0.001$, and Auditory-only conditions, $t(23) = -5.574$, $p < 0.001$. The Auditory-only conditions led to similar perceived mental demand compared to the Visual/Auditory displays, $t(23) = 0.587$, $p = 0.560$.

Physical workload

ANOVA results showed no effect of Orientation on physical workload, $F(1,23) = 0.43$, $p = 0.519$, $\eta_p^2 = 0.018$ (Figure 5.4). 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-only conditions led to significantly higher perceived physical workload compared to Visual/Auditory, $t(23) = 3.580$, $p = 0.002$, and Auditory-only conditions, $t(23) = -2.904$, $p = 0.011$. The Auditory-only conditions led to similar perceived physical demand compared to the Visual/Auditory displays, $t(23) = -0.314$, $p = 0.755$.

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 5.4). 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-only conditions led to significantly higher perceived temporal workload compared to Visual/Auditory, $t(23) = 4.225$, $p < 0.001$, and Auditory-only conditions, $t(23) = -3.386$, $p = 0.003$. The Auditory-only conditions led to similar perceived temporal demand compared to the Visual/Auditory displays, $t(23) = 0.726$, $p = 0.353$.

Performance

ANOVA results showed no significant effect of Orientation on performance, $F(1,23) = 2.878$, $p = 0.103$, $\eta_p^2 = 0.111$ (Figure 5.4). The Display factor showed a significant effect on performance, $F(2,46) = 12.67$, $p < 0.001$, $\eta_p^2 = 0.355$. There were no statistically significant interactions between Orientation and Display, $F(2,46) = 2.268$, $p = 0.115$, $\eta_p^2 = 0.090$. Paired t-tests showed significantly better perceived performance for the Auditory-only conditions compared to the Visual-only 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-only conditions, $t(23) = 2.655$, $p = 0.022$.

Effort

ANOVA results showed significant main effects for Orientation, $F(1,23) = 6.876$, $p = 0.015$, $\eta_p^2 = 0.230$, indicating horizontal movement planes were more effortful to use than vertical movement planes (Figure 5.4). The Display factor also showed a significant main effect on perceived effort, $F(2,46) = 7.708$, $p = 0.001$, $\eta_p^2 = 0.251$. There were no statistically significant interactions between Orientation and Display, $F(2,46) = 0.746$, $p = 0.48$, $\eta_p^2 = 0.031$. Paired samples t-tests showed significantly higher effort for the Visual-only conditions compared to Visual/Auditory, $t(23) = 3.11$, $p = 0.010$, and Auditory-only conditions, $t(23) = -2.941$, $p = 0.010$. Auditory-only conditions and Visual/Auditory conditions were statistically equivalent, $t(23) = 0.120$, $p = 0.905$.

Frustration

ANOVA results showed no significant effect on frustration, $F(1,23) = 4.044$, $p = 0.056$, $\eta_p^2 = 0.150$ (Figure 5.4). The Display factor showed a significant effect on frustration, $F(2,46) = 2.375$, $p = 0.104$, $\eta_p^2 = 0.094$. There were no statistically significant interactions between Orientation and Display, $F(2,46) = 1.745$, $p = 0.186$, $\eta_p^2 = 0.071$. But the pairwise t-tests showed no significant differences between Visual/Auditory and Visual-only prototypes, $t(23) = 2.370$, $p = 0.066$, the Visual/Auditory and Auditory-only, $t(23) = 0.414$, $p = 0.681$, or the Visual-only and Auditory-only system, $t(23) = -1.638$, $p = 0.216$.

Overall workload

The overall workload scale, which was not included in Experiment 1 and 2, 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 answer which among each pair of subscales contributed more to their workload. ANOVA results showed a significant effect of Orientation on overall workload, $F(1,23) = 9.884$, $p = 0.005$, $\eta_p^2 = 0.301$ (Figure 5.5). The Display factor also showed a significant effect on overall workload, $F(2,46) = 15.05$, $p < 0.001$, $\eta_p^2 = 0.396$. There were no statistically significant interactions between Orientation and Display, $F(2,46) = 2.175$, $p = 0.125$, $\eta_p^2 = 0.086$. Pairwise t-tests showed the Visual-only conditions led to significantly higher perceived overall workload compared to Visual/Auditory, $t(23) = 5.045$, $p < 0.001$, and Auditory-only conditions, $t(23) = -5.037$, $p < 0.001$. The Auditory-only conditions led to similar perceived overall workload compared to the Visual/Auditory displays, $t(23) = -0.076$, $p = 0.939$.

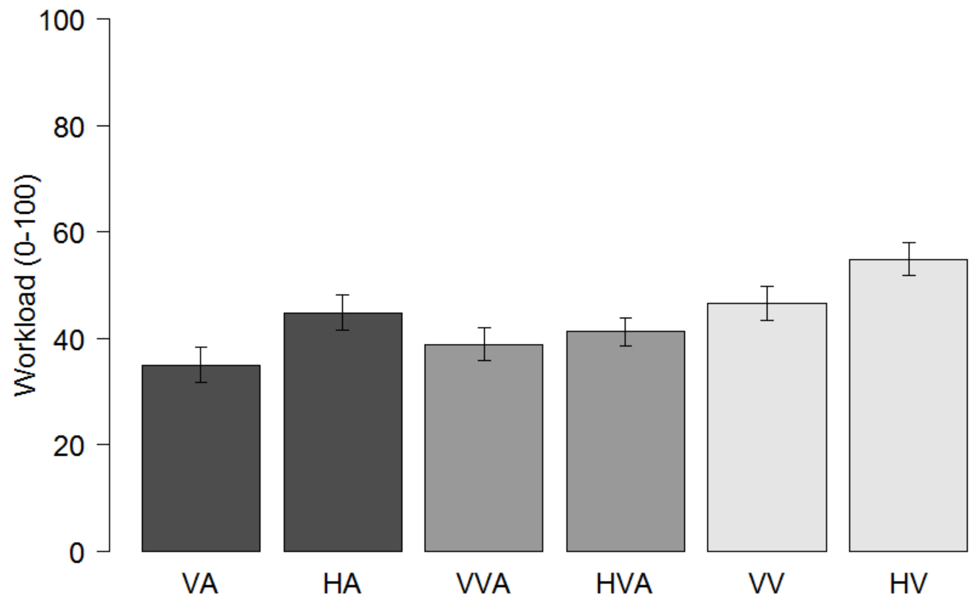

Figure 5.5 overall workload scores for each condition.

5.4 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 impact. 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, lower standard deviation of lane position, and shorter average following distance. 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 driving for visual processing resources. The benefits of the auditory displays on driving were not observed in Experiments 1 and 2 because the other factors had larger effects, leading to greater variance for the Auditory Display when collapsed across other factors. This experiment saw very small effects, in general, for the Orientation factor, making it more possible to identify effects of Displays. This improvement in lane departures associated with auditory-only displays is also 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, Williamson, and Brewster, 2017). One explanation for the inconsistency is that the driving task in the study from Shakeri, Williamson, and Brewster required the driver to drive in a straight line, whereas the driving scenario from Experiment 3 of this dissertation required adapting to changes in speed from a lead vehicle and also adapting to curves on the road. The added difficulty in the driving scenario from Experiment 3 potentially makes driving performance metrics more sensitive to the differences 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 Experiment 3.

Regarding eye glance behavior, the display modality factor had significant impacts on the frequency of short, medium, and long eye off-road eye glances. Conditions with auditory displays resulted in feweroff-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 results from Experiments 1 and 2, and also 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, Williamson and Brewster (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 selectionscompared 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 measure. This result is consistent with findings from May, Gable, and Walker (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, Gable, and Walker (2014), which is the same result observed in Experiment 3.

Workload results showed 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. The visual display factor was associated with poorer perceived performance.

Hypothesis 1 stated that auditory-only displays would lead to slower and less accurate target selections. These predictions were logical extensions from observations in Experiment 2 that conditions with auditory displays led to slower selection times and lower accuracy. In Experiment 3, results showed slower selection times for auditory-only conditions compared to visual-auditory or visual-only conditions. However, the results also showed auditory-only 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 are getting 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 items. The lack of information about fine-grain spatial resolution in the auditory displays supports possible explanation for lower accuracy observed in Experiment 2 for the in-air gesture prototypes – participants could be relying on the auditory displays, even when the visual displays are available, but missing the important information about how centered they are within a menu item. The same effect did not reach statistical significance in Experiment 3, so it seems that if the addition of auditory displays does degrade accuracy, the effect size is relatively small.

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. This hypothesis was also posited and supported in Experiments 1 and 2. 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). Since Experiments 1 and 2 showed the prototypes with visual displays led to greater numbers of off-road glances, then if the same holds true for this experiment 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 showed 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: first, that the workload felt by the participants when using the auditory-only display was less 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 visual display factor, meaning that the addition of a visual display did not 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. In fact, numerically, the physical demand was higher for prototypes with the horizontal control orientation, although the effect did not reach statistical significance. The assumption that the horizontal control orientation would be physically easier 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 driver behavior, because drivers raised their arms while keeping their elbows low, leading to relatively lower physical demand. 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 along the z-axis (forward and backward) were lower than movements on the x-axis (left and right) and movements on the y-axis (up and down). This hypothesis was also consistent with some observations made during Experiment 1 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 led to a ceiling effect where all participants were performing similarly well in making accurate selections. The observation from the first experiment 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, with menu items in the upper right corner of the lower left quadrant leading to greater accuracy. It is possible that if the menu items were smaller than 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 data. This hypothesis was likely unmet for a similar reason, a ceiling effect of physical demand. 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 low (large targets) leading to fast selection times and less movement and more recovery time for drivers.

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. These results demonstrate even more clearly than Experiments 1 and 2, the benefits of auditory displays by taking the extreme case, of auditory-only menus and revealing that they can perform as well or better than similar air gesture control systems with visual displays. The biggest drawback of auditory displays is that they appear to require a longer time to use to make secondary task selections. This should come as no surprise, given the greater bandwidth afforded by the visual modality in comparison to auditory displays. As supported by results from this Experiment, as well as Experiments 1 and 2, the addition of auditory displays presents a tradeoff between driving safety and efficient secondary task performance.

Chapter 6

6 Experiment 4 – Measuring the influence of auditory displays on target acquisitions

6.1 Introduction

Fundamentally, the use of in-air gesture prototypes can be deconstructed into two primary subtasks: search and target acquisition. Through the first three studies, I have investigated the impacts of in-air gesture control use in vehicles. However, a clearer understanding of how searching and target acquisition works could lead to improvements in the selection times and selection accuracy, and potentially reduce visual demands and driver workload. In particular, I measured target acquisition performance in a traditional discrete target acquisition task to compare movements using visual feedback, auditory feedback, and visual/auditory feedback. I compared two different sonification strategies: one discrete (“beep” when over target), and one continuous (continuous tone that increases in loudness and pitch as hand approaches target). Comparing each of these sonification strategies showed the influence of auditory feedback, visual feedback, and combinations of displays on the relationship between index of difficulty (ID) and movement time, accuracy, and overall movement performance. I am also interested in answering several other basic questions about using auditory displays to convey information about in-air target acquisition. How does movement time and accuracy compare between movements made with visual displays and auditory displays? What type of display best facilitates movement performance? These questions are the motivation for my review of past research in movement science and my experiment to investigate the impacts of auditory display on rapid aimed movements.

6.1.1 Literature Review

Fitts's Law

Any discussion of target acquisition should begin with an introduction to Fitts's Law (Fitts, 1954; Fitts & Peterson, 1964). In Fitts's original experiment, he presented participants with a metal plate, on which there were two rectangular targets (Figure 6.1). Participants were asked to tap one target, then the other, and move back and forth tapping each of the targets as quickly as possible while maintaining high accuracy.

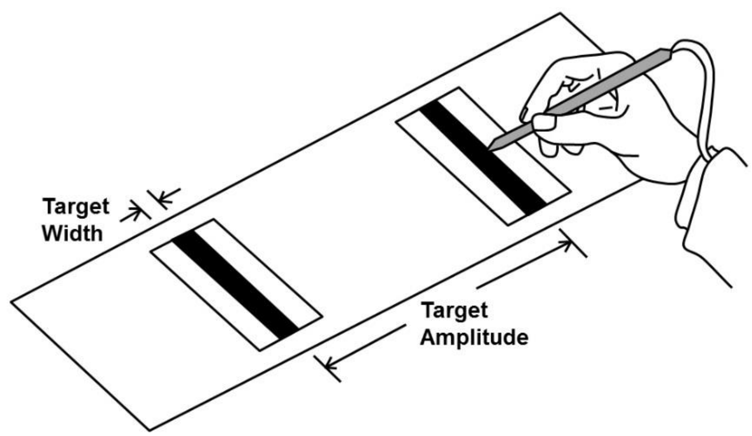

Figure 6.1 Fitts's original reciprocal tapping task setup (Fitts, 1954; Image from MacKenzie, 1992 [copyright permission granted]).

In Fitts's tasks, the size of the targets and the distance between the targets varied for each experimental condition. The finding was that movement difficulty, also called *index of difficulty* (ID), as defined in (1), MacKenzie's (1992) more widely used modification to Fitts's original formula has a positive linear relationship with movement time (MT), as described in (2). In equation 1, A is the distance to the target, or amplitude, and W is target width. In equation 2, *a* and *b* are empirically determined constants arrived at through regression analysis.

$$ID = log_2\left(\frac{A}{W} + 1\right), \quad (1)$$

$$MT = a + b * ID, \quad (2)$$

The ID is determined by the ratio of the target distance to the target width. In other words, as distance to the target increases, and/or as the target size decreases, the ID increases. Since ID is proportional to MT, it is possible to predict that smaller targets take longer to select, and more distant targets will also take longer to select. This may not come as a surprise, as this finding is consistent with many of our intuitions and daily experiences. The profundity of this finding and the reason it is one of the most frequently used models of human movement, is that the same relationship is seen across many populations (Sugden, 1980; Wade, Newell & Wallace, 1978: Brogmus, 1991), modes (MacKenzie, 1992; Kerr, 1973: Drury, 1975), and ranges of movement (Langolf, Chaffin & Foulke, 1976). Fitts's Law is useful to Human Factors researchers because as long as the distance to a target and the size of a target is known, it can predict the amount of time it will take for a person to move to a target, which can inform more usable designs of interactive elements, such as nuclear power plant controls, websites, and in-vehicle information systems.

Researchers have made efforts to translate findings from basic research in movement science, like that from Fitts, to practicable design guidelines for technology interfaces. In an attempt to standardize kinematic evaluations in the field of human-computer interaction, Soukoreff and MacKenzie (2004) advocated for the use of throughput as a measurement of movement performance. Throughput (TP), as defined below (3), states that TP represents the average *performance index* (IP) across all movements, for all participants, where IP (4) is defined by the ratio of *effective index of difficulty* (ID_e)(5) over movement time (MT), in which W_e is defined as a constant multiplied standard deviation of endpoint position along the axis of movement (6).

$$TP = \frac{1}{Y}\sum_{i=1}^{y}\left(\frac{1}{x}\sum_{j=1}^{x}\frac{ID_{e_{ij}}}{MT_{ij}}\right), \quad (3)$$

$$IP = \frac{ID_e}{MT}, \quad (4)$$

$$ID_e = log_2\left(\frac{D}{W_e} + 1\right), \quad (5)$$

$$W_e = 4.133\sigma, \quad (6)$$

Throughput allows us to calculate movement performance by incorporating both movement time and accuracy (Soukoreff & MacKenzie, 2004). The units are bits per second (or bps). Higher bits per second indicates greater movement bandwidth, meaning the it is a more efficient movement, when considering both accuracy and speed.

Kinematic Features of Movement

There are certain features of movement that show consistent relationships with the task demands, i.e., distance to target, target size, and index of difficulty (ID) of the movement. Movement distance impacts the initial ballistic phase of movement. If the distance is greater, the acceleration phase is shorter, and peak acceleration and velocity is higher

(Heath, Hodges, Chua & Elliot, 1998; MacKenzie, Marteniuk, Dugas, Liske & Eickmeier, 1987). Target size also impacts kinematic features. As the target size decreases, the peak velocity decreases, and the duration following peak velocity lengthens (MacKenzie, et al., 1987; Langhoff, Corcoran, Sim, Weinhold & Glover, 1976; Elliot, et al., 1991; Chua & Elliot, 1993). Index of difficulty (ID) also impacts the velocity profile. Generally, movements with higher IDs result in velocity profiles with long tails on the right side, as a result of a longer deceleration and slower approach to targets (Mottet & Bootsma, 1999; Bootsma, Fernandez & Mottet, 2004) (Figure 6.2).

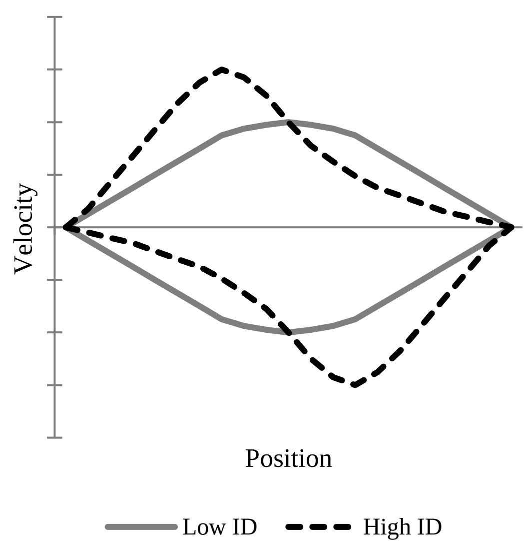

Figure 6.2 velocity profiles for low difficulty and high difficulty movements in a reciprocal tapping task.

The asymmetry of a velocity profile has been used by many researchers as a measure of online regulation of movement (Elliot & Hansen, 2010). That is, the more skewed the profile is, the more the movement is being regulated through incorporation of some information about limb and target position.

Models of motor control

While Fitts's seminal papers provide a robust model of human movement, they do not explain the mechanisms underlying the capability of the visuo-motor system to complete these tasks. Dating back to Woodworth in 1899, researchers have investigated speed-tradeoffs, motor control, and motor learning in goal-directed movements (Woodworth, 1899). Woodworth is credited for his discovery that goal-directed manual aiming is comprised of two movement phases: an initial, *ballistic* phase, and a corrective, *control* phase. In the ballistic phase, the hand is unguided and moves relatively fast. The purpose of the ballistic phase is to move the hand within the vicinity of the target. Over the second half of the movement, the control phase, the hand moves relatively slowly and makes more corrective movements to accurately locate the target.

Many researchers have followed in Woodworth's footsteps. Among the most active areas of research following in Woodworth's wake surrounds the question about which model of motor control best explains the relationship between movement time and index of difficulty. There were three major models that were proposed to represent the way people move their limbs in the control phase of the corrective movements. The first model was the *iterative corrections model* (Keele, 1968; Crossman & Goodeve, 1983). The iterative corrections model states that goal-directed movements are actually comprised of several smaller discrete movements, each of which covers a fixed proportion of the overall distance. Modern equipment has allowed for the measurement of velocity over the duration of movements, which revealed that movements are not made of discrete submovements (Rosenbaum, 2009). The *impulse variability model* says that movements are actually a result of a single impulse that propels the limb over the entire distance (Schmidt, Zelaznik, & Frank, 1978). The second half of the movement in this model is suggested to be simply passive movement derived from the initial impulse. According to the impulse variability model, the force and duration of the impulse are the factors that determine where the limb will stop. This model does a good job describing rapid movements (Schmidt, Zelaznik, Hawkins & Frank, 1979) and movements made without visual feedback (Wallace & Newell, 1983). The *stochastic optimized submovement model* (Meyer, Abrams, Kornblum, Wright & Smith, 1988) is a hybrid of the iterative corrections model and the impulse variability model. It assumes that movements are comprised of an initial impulse which propels the limb towards a target. If the limb lands on the target then the movement is completed. If the limb misses the target, a second impulse moves the limb, and so on until the limb lands on the target. This is the most recent, and most widely accepted model of motor-control in goal-directed movements.

Vision and Goal-directed Movement

Vision has long been assumed to play an integral role in goal-directed movements. Woodworth (1899) hypothesized that people use visual information about the relative position of their limbs and the target to correct their trajectory. In one experiment, Woodworth (1899) asked participants to perform aiming movements to the beat of a metronome, with different speed settings. In one condition, participants performed the tasks with vision, and in another condition they were asked to close their eyes. At slower metronome speeds, participants had lower error with their eyes open because they were able to correct their movements using visual information. Woodworth found that as the speed of the metronome increased, the error rates converged. At approximately 450 ms, the error rates were the same. Woodworth says this was because the faster speed means there is not enough time to reach a control phase at all. Later experiments from Keele and Posner (1968) found that Woodworth's time of 450 ms was too high because he was including the time to reverse direction in a reciprocal tapping task. Keele and Posner found the minimum time for a movement to incorporate visual information into movement correction was somewhere between 190-260 ms. Later, researchers found that even though visual feedforward information can improve future performance, *online* visual information (information about the current movement) is more important (Zelaznik, Hawkins & Kisselburgh, 1983). Here, information means knowledge of limb position as well as target location. Although it was temporarily in doubt (Carson, 1981), both target and limb positions provide important information for goal-directed movement (Elliot, Helsen & Chua, 2001).

Carlton (1979) and Chua and Elliot (1993) found that initial movements in target acquisition tasks would frequently undershoot the target. Interestingly, this is contrary to the assumptions of the optimized submovement model, which states that endpoints should vary around the target center in a normal distribution. Whether or not the cursor is visible to the mover, the variability in the initial movement was unchanged. Even though the number of secondary, corrective, movements was the same across visual and non-visual conditions, the non-visual condition led to greater endpoint variability (Meyer et al., 1988). Meyer et al. suggested that this was because visual information was not available and participants were forced to use less precise kinesthetic information, rather than visual information. Researchers later showed when participants have visual information, they spend more real and proportionate time in the control phase of the movement in traditional goal-directed movement tasks (Elliot, Garson, Goodman & Chua, 1991; Carson, Goodman, Chua & Elliot, 1993), as well as in computer aiming tasks (Chua & Elliot, 1993; Elliot, Lyons & Dyson, 1997). In addition, the visual feedback did not affect kinematic markers like peak velocity and time to peak velocity. Elliot et al. (1991) suggested that the extra time was a result of visual processing, done with the purpose of reducing target-aiming error. That same study showed that the availability of visual information reduces aiming errors by 300% in the accuracy-emphasized condition. In fact, the number of discrete discontinuities in the acceleration profile, i.e., the number of corrections, was similar between conditions with and without visual information. The implication is that visual information does not increase the number of corrections but makes corrections to movement trajectory more efficient. This led Elliot and colleagues to hypothesize that maybe corrections as a result of visual information are continuous, rather than discrete, as previous models have suggested.

In summary, goal-directed movements are generally thought being comprised of two parts, a stereotyped first phase, following a single impulse, and a second phase which is responsible for trajectory corrections. Visual information reduces endpoint error in goal-directed target acquisition, primarily by regulating corrective movements in the second phase of movement, as the limb approaches the target. Visual information has also been shown to be more important for aiming tasks that demand high accuracy.

Auditory Information and Goal-directed Movement

While visual information is closely linked to goal-directed movement performance and has been subject to a lot of investigation, the relationship between auditory information and goal-directed movement performance is less clear. Auditory information has been shown to improve target acquisition in 3D virtual environments (Pierno, Caria & Castiello, 2004; Pierno, Caria, Glover & Castiello, 2005; Zahariev & MacKenzie, 2008). However, these studies asked participants to localize the source of a sound in 2D or 3D space around themselves. Our interest is to learn about using auditory displays to measure how close their hand/cursor is to a target. Hatfield, Wyatt, and Shea (2010) found that adding auditory contact cues reduced movement times for reciprocal tapping tasks and effectively reduced the movement difficulty (ID). My goal is to convey information about hand position relative to target, rather than the relative position of the participant's head to the target. Sonification of kinematic features of movements in sporting events, has been demonstrated to improve recognition of kinematic features (Effenberg, 2001) and reproduction of movements (Effenberg & Mechling, 2003). However, these sonifications are conveying information about kinematic features, e.g., maximum velocity, height, etc., whereas I am trying to convey distance between the current position of the hand and the intended target position. As far as I know, there is little or no research investigating the potential of movement sonification on improving target acquisition in this way.

Goal-directed Movement, Visual Information, and In-vehicle Gesture Controls

The abundance of research on target acquisition has been conducted on participants with full vision of their limbs and target locations and some research has deprived participants of some or all of that visual information. However, very little research has investigated the influence of auditory information on these same kinematic features and measures of movement performance. For my in-vehicle gesture controls to improve driving safety, I need to facilitate fast, accurate movement while minimizing the necessity of visual information. I substituted an auditory display for a visual one. If I can successfully convey information about the relative position of the driver's limb and the menu items, then I should see similar movement performance for individuals using systems with visual displays only and those using auditory displays only. As far as I am aware, there has been little investigation into the topic of auditory-aided eyes-free movement performance. My work in this area is motivated by my pursuit of suitable auditory displays for in-vehicle use. As such, my priority is not only to better understand the relationship between auditory information and movement performance, but also to understand its impacts on workload, and ultimately its impacts on driving performance as well. My proposed experiment will not cover the topic of auditory displays in-vehicles in an all-encompassing way, but is intended to determine the relative performance benefits of an auditory display relative to a visual display, generally, and also to give early direction on which types of auditory displays will be most appropriately applied to an in-vehicle control system.

Hatfield, Wyatt, and Shea (2010) already found that the addition of a simple tone when participants enter a target zone, shortens movement times and reduces skew in velocity profiles. Their concern was about the lack of auditory feedback inherent in computer-based Fitts's tapping tasks may be affecting the fit of the models to the data. I would like to investigate the impacts of short auditory tones on selection accuracy and effective target width. However, the major thrust of my research in this area surrounds the question, can auditory information be substituted for visual information in the control of movements? How does a continuous sonification of the distance between the participants' hand and the target position influence selection time, selection accuracy, and throughput? As far as I know, this experiment would be the first experiment to validate the impact Hatfield et al. observed via in-air gestures. Furthermore, this experiment may be the first to investigate the impact of a continuous sonification of hand-to-target distance on movement performance in a target acquisition task.

Some studies have investigated aimed movement performance using air gesture controls (e.g., Grossman & Balakrishnan, 2004; Cockburn et al., 2011) and even explored the concept of eyes-free aimed movements (Cockburn et al., 2011) using only kinesthetic information. Other studies have examined the impact of auditory displays on target acquisition performance (Hatfield, Wyatt & Shea, 2010; de Grosbois, Heath & Tremblay, 2015). However, to our knowledge there is little to no existing literature exploring the utility of auditory displays in conjunction with air gesture controls in aiding target acquisition tasks. Most existing literature surrounding the topic of auditory displays and air gestures have focused on target localization, i.e., finding the point of origin of a sound in space (e.g., Soukoreff & MacKenzie, 2004; Pierno, Caria & Castiello, 2004; Pierno et al., 2005; Zahariev & Mackenzie, 2008; Marentakis & Brewster, 2004).

We conducted an experiment to learn how auditory displays affect aimed movement performance using air gesture controls. We made comparisons between two sonification strategies: (1) a discrete auditory display – playing a sound whenever the user is on the target and (2) a continuous auditory display – playing sound continuously from the start of the movement until selection, and playing a discrete sound when the user is on target. We also made comparisons among auditory-only, visual-only, and visual-auditory displays, as well as a control condition for which there was no visual or auditory display.

Soukoreff and MacKenzie (2004) wrote a paper outlining several guidelines which supported the ISO 9241-9 standards for the evaluation of pointing devices in human-computer interaction. In keeping with standard evaluation of pointing devices, we followed each of those standards as much as possible. This is our justification for (1) our use of the Shannon formulation of index of difficulty, (2) our range of movement difficulties, (3) our adjustments for selection accuracy, (4) and our calculation of throughput.

6.2 Hypotheses

6.2.1 Movement Time

H1: The slowest movement times will come from the continuous auditory display without visuals. The second slowest conditions will be the continuous auditory display with visuals, and the visual display with no auditory display. The discrete auditory display with visuals will be faster than all previously mentioned conditions. Next fastest will be the discrete auditory display condition without visuals. Finally, the fastest condition should be the no-display condition. For all conditions, ID is expected to increase movement times.

I assume that a continuous auditory display will result in very slow movements because the auditory display will serve as the only non-kinesthetic source of information. Processing distance data via an auditory display may be slower and less intuitive than a visual display, and will require slower movements to react and integrate information from the auditory display, whereas visual information is readily integrated into trajectory corrections (Elliot, Helsen & Chua, 2001). It is assumed that for conditions with less information, such as the discrete auditory display condition, and the no-display condition, that participants will satisfice with the little available information, which may result in a shift in emphasis towards speed rather than accuracy. Where more information is available, visual information will be preferred for its simpler, faster integration into the motor system. Continuous auditory displays, however, may provide sufficient information to make accurate selections. Continuous auditory information, because it is less familiar and less intuitive – among perhaps other reasons – will take more time to interpret and translate into online movement corrections. I anticipate that slower movements will be necessary to allow for the slower processing and integration of continuous auditory information.

6.2.2 Movement Accuracy

H2: Conditions in which there is relatively little information about the difference between target and hand positions, will shift towards a speed-emphasized strategy in the speed-accuracy tradeoff associated with goal-directed movements. Conditions such as the no-display condition and the discrete auditory display condition provide little or no information about the relative position of the hand and target, which may be insufficient to make accurate movements. As a result, those two conditions should result in fewer correct target selections and greater variance in movement endpoints compared to other conditions. Conditions with visual displays will result in higher accuracy. Auditory displays, when paired with visual displays are expected to have little or no measurable impact on accuracy results. The continuous auditory display, with no visual display, is expected to be as accurate as the visual displays. Results from my previous studies suggest that the addition of discrete auditory displays has no influence on accuracy for visual-auditory displays. It is also my assumption that information from auditory displays will be sufficient to make accurate movements.

Effective target widths will be measured and plotted to allow for visual comparisons of the spread of selection points for each target. Effective width will be plotted against actual target width. This can reveal the relationship between the display type, and ID and size of targets that can be selected in fast goal-directed movements.

6.2.3 Movement Performance

H3: Throughput is simply a calculation derived from the movement time and accuracy. So, the hypotheses follow logically from the hypotheses stated in the movement time and movement accuracy sections. The visual-auditory display with discrete auditory feedback will result in best throughput performance, followed by the visual-only condition, and the visual-auditory display with the continuous feedback. The lowest performing conditions will be the auditory-only conditions, in this order: continuous auditory display, discrete auditory display, and no-display condition. Throughput will be lower for movements of higher IDs. Similar patterns will be observed when accuracy is plotted against movement time.

6.3 Methods

6.3.1 Participants

A total of 24 undergraduate psychology students were recruited to complete our study (Table 6.1). All participants were given course credit as compensation for their participation. All participants had normal or corrected-to-normal vision and hearing. Only one person reported having experience using a LEAP Motion before.

Table 6.1 participant demographics

Age (yrs)	**Gender**	**Handedness**
Mean=19.75	Males:14	Right:21
SD=1.96	Females:10	Left:3

6.3.2 Apparatus

We used a LEAP Motion to track hand position and we used Pure Data – an open source graphical programming language – to design the visual and auditory displays for our target selection task (Figure 6.3). The LEAP Motion was mounted at a 45 degree tilt on the table surface. This was done to avoid fatigue because the sensor performed better if participants' hands were in a parallel plane to the sensor plane, which would require participants to raise their elbows above the table height. As the participant moves their hand above the sensor, the position data are relayed to the Pure Data patch which displays a cursor tracking the x position (no y-position data was displayed or recorded) of the participant's hand on a 19" monitor. Pure Data also recorded (at 100 Hz) all position data and time data, allowing us to analyze the movement performance. There was no gain associated with movements along the x-axis. That is, if a participant moved one inch to the right, the cursor moved one inch to the right on the screen. The task required participants to select the targets. In order to maintain consistency with the previous experiments, we decided to use the same selection gesture, i.e., opening hand to reveal five fingers. This gesture works best given the constraints of the sensor because counting fingers is relatively easy. Counting to five fingers has added the benefit that it is reduces the probability of false positives that arise when the sensor loses track of the position temporary, in which case it counts zero visible fingers.

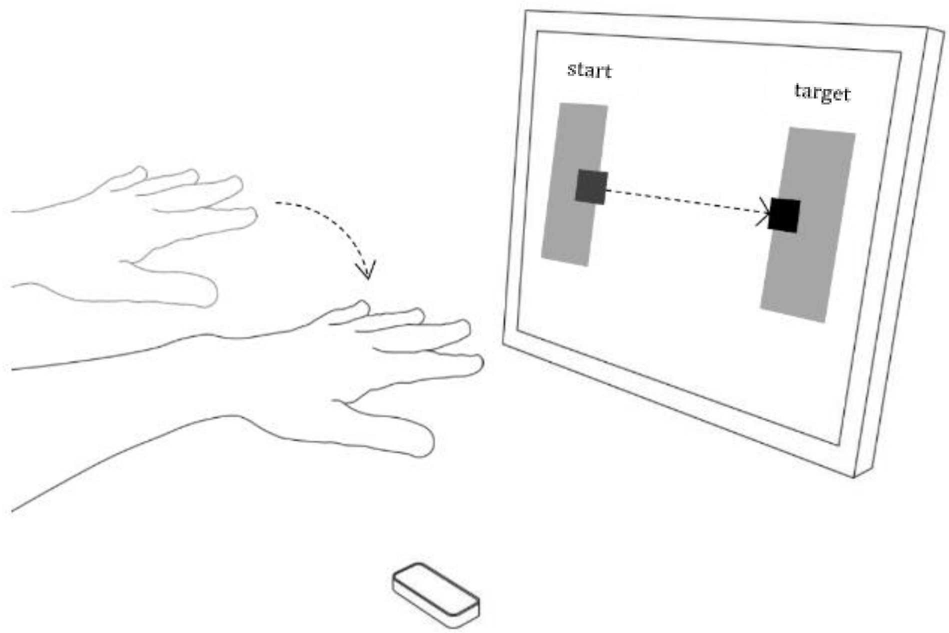

Figure 6.3 Experimental apparatus with LEAP motion and monitor and visual display.

6.3.3 Visual design

The visual display is comprised of a start box, a target box, and a cursor (Figure 6.3). The start box was always located in a fixed position on the left of the screen, labeled “start”, and would appear green. When the cursor, a gray box, entered the start box, the color changed to red. Targets were always located somewhere to the right of the start box at varying distances. Targets also changed from green to red when the cursor entered the box. For conditions without visual displays, the target boxes disappeared immediately after the participants selected the start button. Previous research has shown that removing visual information at the onset of movement allows a visual memory to enhance movement, and masking for 2 or 10 seconds allows the visual memory to decay (Elliot & Jaeger, 1988; Elliot & Madalena, 1987). Despite evidence that participants would be able to use a decaying visual memory of the target position to aim their movements, we decided to keep leave out the masking because the way we controlled the starting point was by allowing participants to select the start button. Once the participant starts a trial, there is nothing to prevent the participant from moving toward the target during the masked period. For this reason, we decided to remove the visual information at the start of the trial. This choice means it is possible that participants could perform better than expected for conditions without visual information.

6.3.4 Sound design

Auditory displays were designed to inform users about the relative position between the cursor and the target position, with the intention that participants should be able to use this information to guide their movements toward their target. There were two different sound designs: a *discrete* auditory display and a *continuous* auditory display. The discrete auditory display consisted only of a pink noise that played whenever the cursor was within the target. The continuous auditory display constantly plays a sine wave that increases in frequency as the cursor gets closer to the target. The frequency of the sound started at 440 Hz and doubled to 880 Hz at target position. The frequency increases as a function of the square (x^2) of current fraction of the total distance to the target that the cursor has traveled (Equation 7). The pitch increases one octave from the start to the target position. The continuous auditory display also played a pink noise when the cursor was within the target position. In other words, as the user moves their hand toward the target, the sound increases in pitch and it resolves as the octave doubles and the pink noise is triggered. If the user overshoots the target the pitch will go down as the distance from the target increases, but will increase again as the participant moves back toward the target.

$$Frequency_t = 440 + \left(\sqrt{440} * \left(1 - \left(\frac{|position_t - target\ position|}{target\ position}\right)\right)\right)^2 \quad (7)$$

6.3.5 Experimental Design

This experiment was a full factorial within-subjects design. Each participant completed movement tasks in six conditions (Table 6.2).

Table 6.2 the six conditions including three levels of auditory factor and two levels of visual factor

	Visual	**No Visual**
Discrete Auditory	VAD	AD
Continuous Auditory	VAC	AC
No Auditory	V	Control

Participants completed a total of 48 movements for each of the six conditions. The block of 48 trials for each condition was comprised of four movement difficulties (IDs), including IDs of 2, 3, 4, and 5 bits ~~(Table 6.2).~~ Twelve movements were completed for each difficulty level.

The IDs were selected based on a guideline written by Soukoreff and MacKenzie (2004), which was also referenced in the ISO 9241-9, a standard for the evaluation of pointing devices in human-computer interaction. Their research suggested using a wide range of IDs when evaluating the pointing devices in human computer interaction. In fact, they even suggest a range from 2-8 bits. In a short pilot study using the prototype apparatus, we observed that performance with our gesture system, which can be considered as a type of pointing device, dropped off significantly at higher IDs due to a combination of sensor noise and lack of fine resolution. As a result, we decided not to include IDs higher than five bits. Interestingly, in Experiments 1 and 2, movement IDs were approximately 2 bits on average (for 2x2 grids), which can provide a real-world basis of comparison for results from this experiment. In order to mirror the target acquisition tasks required to use our gesture interface while driving, we required all selections to be made right-handed starting from a fixed position (assuming their hand starts from the steering wheel), with targets always positioned to the right of the starting point. Our goal was also to reflect the distances required for those types of movements so we varied them between 3, 6, and 9 inch distances. The upper limit was constrained by the size of the screen since we kept the movement gain ratio at one-to-one. The target widths (W), required to achieve the desired IDs, are then calculated using equation (1) (Table 6.3).

$$ID = log_2\left(\frac{A}{W} + 1\right), \quad (1)$$

Table 6.3 difficulties, amplitudes, and widths of movements presented in every condition

ID	**A (in.)**	**W (in.)**
2	3	1
2	6	2
2	9	3
3	3	0.43
3	6	0.86
3	9	1.29
4	3	0.2
4	6	0.4
4	9	0.6
5	3	0.1
5	6	0.19
5	9	29

6.3.6 Procedure

Each experimental session consisted of a short demographic questionnaire, a practice period, and a testing period. Each session required participants to complete a total of 288 (12 trials x 4 difficulties x 6 conditions) target acquisition tasks and took less than an hour to complete.

Practice

After providing informed consent and filling out a brief demographic questionnaire, participants were first introduced to the general purpose of the experiment and given five minutes of guided practice during which they were exposed to each of the six conditions. Participants were seated in a chair in front of a computer and a LEAP Motion fixed to the table at a 45 degree angle to avoid fatigue.

Testing

After selecting the start button (open hand = select gesture), a target appeared somewhere to right on the screen. After selecting the target, all visual and auditory displays were removed for a few seconds before the start button appears for the next trial. Participants were allowed to complete the task with their left or right hand but they were asked to not switch hands during the experiment. Participants were encouraged to take breaks between selections or conditions as needed.

6.3.7 Dependent Variables

Time

Movement Time – Time from start of movement until the target selection, or endpoint. Calculation of average movement times included only correct selections, or selections where the endpoint landed within the target box.

Accuracy

Movement Accuracy – Percentage of movement endpoints inside the target area.

Adjusted Error – Distance of movement endpoints from the edge of the nearest target.

Effective Target Width – Standard deviation in endpoint from the target center in the primary direction of movement (Equation 6), effective target width can be calculated to measure selection accuracy.

$$W_e = 4.133\sigma, \tag{6}$$

Movement Performance

Throughput –as defined by Soukoreff and MacKenzie (2004) in (3). This provides a measure of overall movement performance, including both movement accuracy and selection time.

$$TP = \frac{1}{Y}\sum_{i=1}^{y}\left(\frac{1}{x}\sum_{j=1}^{x}\frac{ID_{e_{ij}}}{MT_{ij}}\right), \tag{3}$$

6.3.8 Statistics

Repeated-measures ANOVAs were conducted to identify main effects. Paired t-tests were conducted if a factor showed a significant main effect and there was no three-way interaction between effects. However, if there was a significant three-way interaction, all paired t-tests were conducted. A Holm-Bonferroni correction was used to minimize the number of Type-1 errors for all pairwise comparisons, meaning the critical p-value is 0.05 divided by the number of tests conducted.For the Auditory Display factor (3 levels) the critical p-value is 0.017. For the Difficulty factor (4 levels), the critical p-value is 0.0083.

For selection times, two participants were removed from analysis due to equipment failure.

6.4 Results

6.4.1 Selection time

Repeated measures ANOVA results indicated a main effect for Auditory Display, $F(2,44) = 59.17$, $p < 0.001$, $\eta_p^2 = 0.729$, Visual Display, $F(1,22) = 68.73$, $p < 0.001$, $\eta_p^2 = 0.758$, and Difficulty, $F(3,66) = 71.62$, $p < 0.001$, $\eta_p^2 = 0.765$ (Figure6.4) . There were significant statistical interactions between Visual Displays and Auditory displays, $F(2,44) = 49.7$, $p < 0.001$, $\eta_p^2 = 0.693$, Visual Displays and Difficulty, $F(3,66) = 9.897$, $p < 0.001$, $\eta_p^2 = 0.310$, and Auditory Displays and Difficulty, $F(6,132) = 11.76$, $p < 0.001$, $\eta_p^2 = 0.348$. There was also a significant three-way interaction between Auditory Displays, Visual Displays, and Difficulty, $F(6,132) = 17.59$, $p < 0.001$, $\eta_p^2 = 0.444$. This interaction effect can be observed in Figure 6.4. The steepness of the slope for the auditory-only conditions is much greater (slower selection times) at higher difficulties compared to all other conditions, including conditions with visual-only displays and visual-auditory displays.

Paired comparisons (Table 6.4) showed participants were slower to make selections when using continuous (AC) and discrete (AD) auditory displays compared to conditions with visual displays and the control condition (Figure 6.4). The AD condition led to slower selection times than the AC condition, $t(22) = 5.278$, $p < 0.001$, Control condition, $t(22) = -13.919$, $p < 0.001$, V condition, $t(22) = -14.000$, $p < 0.001$, VAC, $t(22) = -14.408$, $p < 0.001$, and VAD condition, $t(22) = -13.270$, p < 0.001. The AC condition led to slower task completion times than Control condition, $t(22) = - 12.502$, $p < 0.001$, V condition, $t(22) = -11.388$, $p < 0.001$, VAC condition, $t(22) = -14.801$, $p < 0.001$, and VAD condition, $t(22) = -14.236$, $p < 0.001$. The VAC condition led to similar selection times as the V condition, $t(22) = -2.654$, $p = 0.047$. Finally, the VAD condition similar selection times as the VAC condition, $t(22) = 3.272$, $p = 0.009$, and V condition, $t(22) = 0.084$, $p = 1.000$.

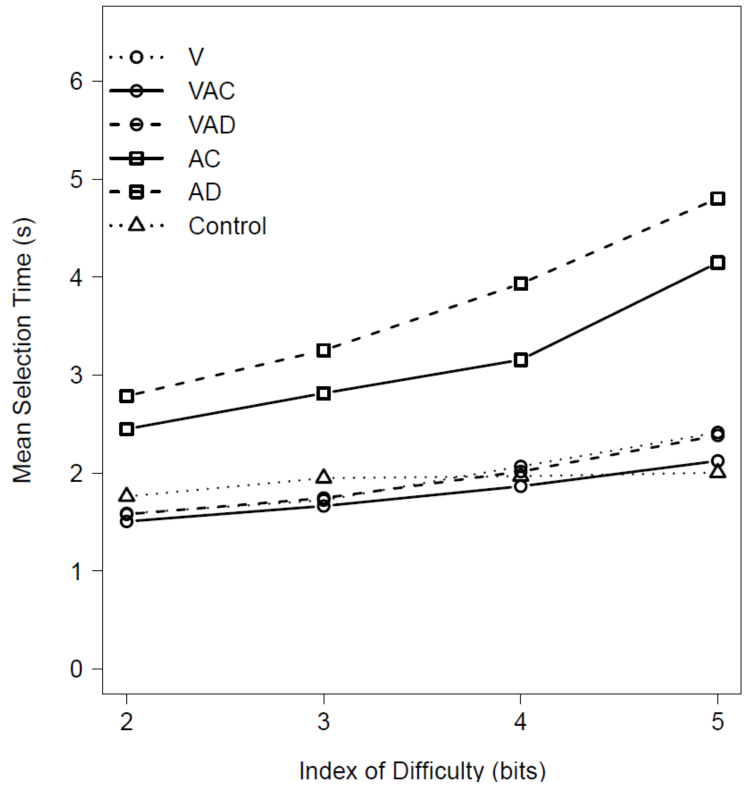

Figure 6.4 Average selection times for each condition across difficulty levels.

Table 6.4 p-values for paired comparisons of average selection times

	AC	**AD**	**Control**	**V**	**VAC**
AD	<0.001*	--	--	--	--
Control	<0.001*	<0.001*	--	--	--
V	<0.001*	<0.001*	1.000	--	--
VAC	<0.001*	<0.001*	1.000	0.047	--
VAD	<0.001*	<0.001*	1.000	1.000	0.009

6.4.2 Selection accuracy

Error

Repeated measures ANOVA results showed a significant main effect for Visual Display on error, $F(1,23) = 56.53$, $p < 0.001$, $\eta_p^2 = 0.711$, and Auditory Display, $F(2,46) = 23.49$, $p < 0.001$, $\eta_p^2 = 0.505$, and Difficulty, $F(3,69) = 54.29$, $p < 0.001$, $\eta_p^2 = 0.702$ (Figure 6.4). All interactions were significant, including: Auditory and Visual Display, $F(2,46) = 28.59$, $p < 0.001$, $\eta_p^2 = 0.554$, Visual and Difficulty, $F(3,69) = 19.25$, $p < 0.001$, $\eta_p^2 = 0.456$, Auditory and Difficulty, $F(6,138) = 7.34$, $p < 0.001$, $\eta_p^2 = 0.242$, and Visual, Auditory display, and Difficulty, $F(6,138) = 7.162$, $p < 0.001$, $\eta_p^2 = 0.237$. Paired comparisons (Table 6.5) showed that participants' selection error, defined by the absolute value of the distance between the final cursor position and the closest edge of the target, was significantly higher for the control condition (Figure 6.4) compared to all other conditions; AC, $t(23) = 9.225$, $p < 0.001$, AD, $t(23) = 9.177$, $p < 0.001$, V, $t(23) = -11.992$, $p < 0.001$, VAC, $t(23) = -12.139$, $p < 0.001$, and VAD, $t(23) = 11.585$, $p < 0.001$. These tests also revealed that the discrete auditory display (AD) led to significantly higher error compared to all conditions other than the control; V, $t(23) = -6.03$, $p < 0.001$, VAC, $t(23) = -6.343$, $p < 0.001$, VAD, $t(23) = - 5.746$, $p < 0.001$. The AC condition led to significantly higher error compared to all conditions other than AD, and Control; V, $t(23) = -5.105$, $p < 0.001$, VAC, $t(23) = -5.526$, $p < 0.001$, VAD, $t(23) = -4.07$, $p < 0.001$. All other conditions were statistically equivalent.

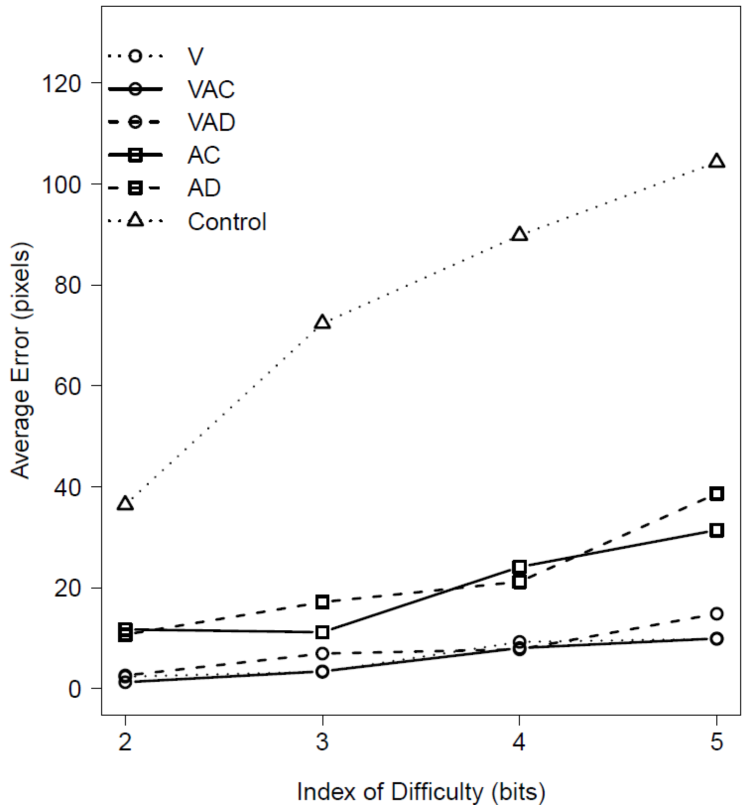

Figure 6.4 average adjusted error for each condition across difficulty levels

Table 6.5 p-values for paired comparisons for selection error across conditions

	AC	AD	Control	V	VAC
AD	0.956	--	--	--	--
Control	<0.001*	<0.001*	--	--	--
V	<0.001*	<0.001*	<0.001*	--	--
VAC	<0.001*	<0.001*	<0.001*	0.957	--
VAD	<0.001*	<0.001*	<0.001*	0.657	0.199

Percent correct

ANOVA results showed a main effect for Visual Display, $F(1,23) = 95.76$, $p < 0.001$, $\eta_p^2 = 0.806$, and Auditory Display, $F(2,46) = 40.6$, $p < 0.001$, $\eta_p^2 = 0.639$, and Difficulty, $F(3,69) = 472.1$, $p < 0.001$, $\eta_p^2 = 0.954$ (Figure 6.5). There was also a significant statistical interaction between the Visual and Auditory Display factors, $F(2,46) = 79.77$, $p < 0.001$, $\eta_p^2 = 0.776$, the Auditory and Difficulty factors, $F(6,138) = 3.575$, $p = 0.003$, $\eta_p^2 = 0.135$. However, there was no significant interaction between the Visual Display and Difficulty factors, $F(3,69) = 0.819$, $p = 0.488$, $\eta_p^2 = 0.034$. There was also a significant three-way interaction between Visual, Auditory, and Difficulty, $F(6,138) = 3.503$, $p = 0.003$, $\eta_p^2 = 0.132$. Figure 6.5 shows that conditions appear to largely be similar with the exception of the control condition which is significantly lower, which can be seen in Table 6.6. Pairwise comparisons showed significantly lower percent correct selections for the control condition, lower than all other conditions; AC, $t(23) = -$

13.009, $p < 0.001$, AD, $t(23) = -14.16$, $p < 0.001$, V, $t(23) = 18.104$, $p < 0.001$, VAC, $t(23) = 17.577$, $p < 0.001$, VAD, $t(23) = 18.368$, $p < 0.001$. The AD condition was also led to lower percent accuracy compared to every condition except for the AC, and Condition; V, $t(23) = 4.558$, $p < 0.001$, VAC, $t(23) = 4.587$, $p < 0.001$, VAD, $t(23) = 4.428$, $p < 0.001$. The VAC condition also led to higher accuracy than the AC condition, $t(23) = 3.638$, $p = 0.003$.

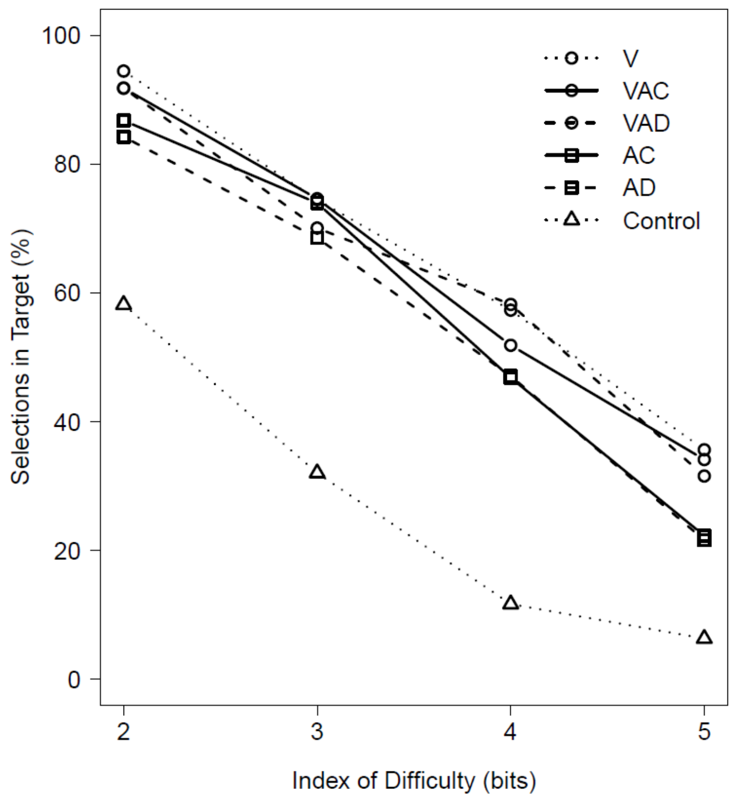

Figure 6.5 average percent correct across conditions for each difficulty.

Table 6.6 p-values for paired comparisons for selection accuracy

	AC	**AD**	**Control**	**V**	**VAC**
AD	1.00	--	--	--	--
Control	<0.001*	<0.001*	--	--	--
V	0.012	<0.001*	<.001*	--	--
VAC	0.003*	<0.001*	<.001*	1.00	--
VAD	0.015	<0.001*	<.001*	1.00	1.00

6.4.3 Undershooting and Overshooting

Among the selections, many were misses. Further investigation into the spatial distribution of those misses could help reveal factors contributing to differences between conditions observed. Comparing undershooting, i.e., missing to the left of the targets (targets always appear to the right of the start box), and overshooting, i.e., missing to the right of targets could give an initial indication about the discrete corrections made during the control phase of target selections. Overshooting is potentially an indication that participants movements passed the target position and were followed by directional changes, which could be contributing factors to time differences observed between auditory-only displays and displays with visual information. The next section investigates the time-to-target. Time-to-target, combined with undershooting/overshooting may indicate whether participants were indeed making more discrete corrections after passing the target. By subtracting the number of overshoots from the number of undershoots, we can get relative difference. This method avoids methodological issues that could arise from participants who made no misses, and should control for differences in frequency of misses associated with higher ID movements (this is the problem with comparing counts of misses). Raw overshoot and undershoot data can be seen in Table 6.7.

Repeated measures ANOVAs showed significant main effects for Auditory Display, $F(2,46) = 25.65$, $p < 0.001$, $\eta_p^2 = 0.517$, Visual Display, $F(1,23) = 33.03$, $p < 0.001$, $\eta_p^2 = 0.579$, and Difficulty, $F(3,69) = 5.636$, $p = 0.002$, $\eta_p^2 = 0.209$. The Auditory Display factor showed a significant interaction with the Visual Display factor, $F(2,46) = 30.31$, $p < 0.001$, $\eta_p^2 = 0.558$, and the Difficulty factor, $F(6,138) = 3.457$, $p = 0.003$, $\eta_p^2 = 0.126$. The Visual Display also showed a significant interaction with Difficulty, $F(3,69) = 8.349$, $p < 0.001$, $\eta_p^2 = 0.258$. There was also a three-way interaction between Auditory, Visual, and Difficulty factors, $F(6,138) = 2.175$, $p = 0.049$, $\eta_p^2 = 0.083$.

Table 6.7 descriptive statistics for over/undershooting across difficulty by condition.

condition	ID	under shoot	over shoot	total miss	% undershoot	% overshoot	ratio (under:over)
Control	2	17	108	125	14%	86%	0.157
AD	2	18	18	36	50%	50%	1.000
AC	2	23	16	39	59%	41%	1.438
VAD	2	13	7	20	65%	35%	1.857
VAC	2	16	7	23	70%	30%	2.286
V	2	13	1	14	93%	7%	13.000
Control	3	46	157	203	23%	77%	0.293
AD	3	57	38	95	60%	40%	1.500
AC	3	43	37	80	54%	46%	1.162
VAD	3	58	27	85	68%	32%	2.148
VAC	3	49	22	71	69%	31%	2.227
V	3	64	12	76	84%	16%	5.333
Control	4	70	197	267	26%	74%	0.355
AD	4	68	87	155	44%	56%	0.782
AC	4	92	60	152	61%	39%	1.533
VAD	4	88	44	132	67%	33%	2.000
VAC	4	100	42	142	70%	30%	2.381
V	4	94	34	128	73%	27%	2.765
Control	5	70	209	279	25%	75%	0.335
AD	5	128	103	231	55%	45%	1.243
AC	5	136	88	224	61%	39%	1.545
VAD	5	145	62	207	70%	30%	2.339
VAC	5	138	61	199	69%	31%	2.262
V	5	135	60	195	69%	31%	2.250

Pairwise comparisons showed significantly more overshooting compared to undershooting (Table 6.8) for the Control condition compared to the AC condition, $t(23) = -6.277$, $p < 0.001$, AD condition, $t(23) = -5.996$, $p < 0.001$, V condition, $t(23) = 8.147$, $p < 0.001$, VAC, $t(23) = 7.556$, $p < 0.001$, and VAD condition, $t(23) = 7.862$, $p < 0.001$. No other pairs were significantly different.

Table 6.8 p-values for pairwise comparisons of undershooting versus overshooting.

	AC	AD	Control	V	VAC
AD	0.439	--	--	--	--
Control	<0.001*	<0.001*	--	--	--
V	<0.704	0.057	<0.001*	--	--
VAC	1.000	0.254	<0.001*	1.000	--
VAD	1.000	0.138	<0.001*	1.000	1.000

Another measure that can illuminate mechanisms underlying time differences in selection times is the difference in time between when a participant first reaches the target position and when they make a final selection. Repeated measures ANOVAs revealed main effects for Visual displays, $F(1,22) = 39.7$, $p < 0.001$, $\eta_p^2 = 0.643$, Auditory Display, $F(2,44) = 39.87$, $p < 0.001$, $\eta_p^2 = 0.644$, and Difficulty, $F(2,44) = 41.63$, $p < 0.001$, $\eta_p^2 = 0.654$ (Figure 6.6). There were also significant two-way interactions between Visual and Auditory Displays, $F(2,44) = 44.4$, $p < 0.001$, $\eta_p^2 = 0.669$, Visual and Difficulty, $F(3,66) = 21.68$, $p < 0.001$, $\eta_p^2 = 0.496$, Auditory Display and Difficulty, $F(6,132) = 12.2$, $p < 0.001$, $\eta_p^2 = 0.357$. In addition, there was a significant three-way interaction between Visual, Auditory displays and Difficulty, $F(6,132) = 18.62$, $p < 0.001$, $\eta_p^2 = 0.458$.

Pairwise comparisons showed longer time difference for the AC condition compared to the Control condition, $t(23) = -8.113$, $p < 0.001$, the Visual-only condition, $t(23) = -5.242$, $p = 0.002$, the Visual with Continuous Auditory display, $t(23) = -7.842$, $p < 0.001$, and the Visual with Discrete Auditory condition, $t(23) = -6.860$, $p < 0.001$. The AD condition also resulted in statistically significantly longer times compared to the Control condition, $t(23) = -8.764$, $p < 0.001$, the V condition, $t(23) = -6.669$, $p < 0.001$, the VAC condition, $t(23) = -7.832$, $p < 0.001$, and the VAD condition, $t(23) = -6.703$, $p < 0.001$. There were no other statistically significant differences (Table 6.9).

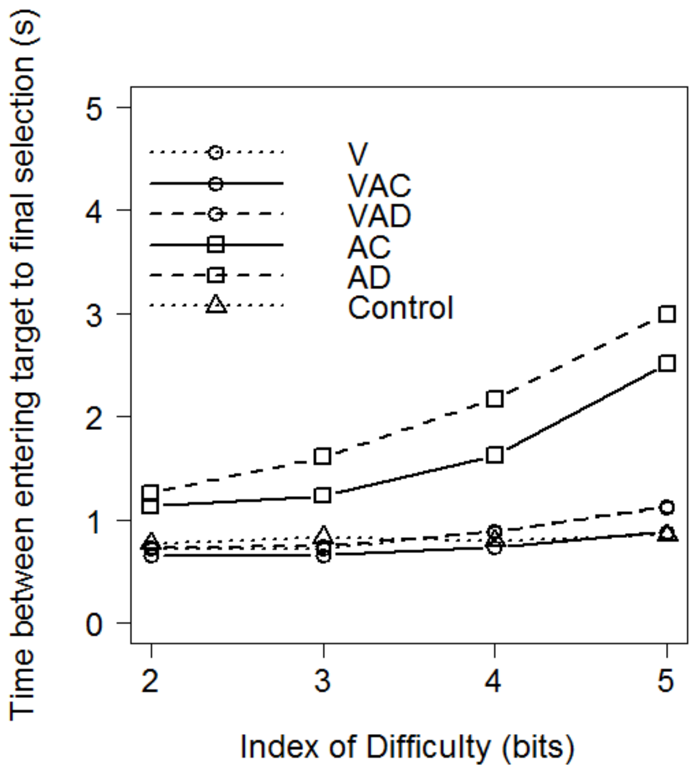

Figure 6.6 Time differences from entering target to making final selection across condition by difficulty.

Table 6.9 p-values for time difference between entering target and final selection by condition.

	AC	AD	Control	V	VAC
AD	0.021	--	--	--	--
Control	<0.001*	<0.001*	--	--	--
V	<0.001*	<0.001*	1.000	--	--
VAC	<0.001*	<0.001*	1.000	0.126	--
VAD	<0.001*	<0.001*	1.000	1.000	0.021

6.4.4 Throughput

Throughput is a calculation that accounts for both the accuracy of the movement – difference between endpoint position and the center of the target – and movement time. This provides a measure of overall movement performance by information conveyed in bits per second. Repeated measures ANOVA results showed no significant effect for Auditory Displays on throughput, $F(2,46) = 1.418$, $p = 0.253$, $\eta_p^2 = 0.058$ (Figure 6.6). There was a significant main effect for Visual Displays, $F(1,23) = 278.8$, $p < 0.001$, $\eta_p^2 = 0.924$. There was also a main effect for Difficulty, $F(3,69) = 14.07$, $p < 0.001$, $\eta_p^2 = 0.380$, and a statistical interaction between Visual Display and Difficulty, $F(3,69) = 6.081$, $p < 0.001$, $\eta_p^2 = 0.209$. However, there was no statistical interactions between Visual and Auditory Displays, $F(2,46) = 0.787$, $p = 0.461$, $\eta_p^2 = 0.033$, Auditory Display and Difficulty, $F(6,138) = 1.585$, $p = 0.156$, $\eta_p^2 = 0.064$, or Visual Display, Auditory Display, and Difficulty, $F(6,138) = 1.134$, $p = 0.346$, $\eta_p^2 = 0.047$. Paired comparisons showed that the continuous auditory displays led to greater throughput compared to the discrete auditory displays, $t(23) = 4.517$, $p < 0.001$, as did the no display conditions, $t(23)$

= -3.134, $p = 0.004$. But the continuous auditory displays led to equivalent throughput compared to the no auditory display conditions, $t(23) = 0.830$, $p = 0.407$ (Figure 6.6). Paired t-tests for Difficulty showed 2 bit and 3 bit difficulties were similar, but resulted in higher throughput than 4 and 5 bit difficulties. Highest difficulty, 5 bit IDs led to lower throughput compared to 4 bit, $t(23) = -6.291$, $p < 0.001$, 3 bit, $t(23) = -8.892$, $p < 0.001$, 2 bit $t(23) = -9.900$, $p < 0.001$. The 4 bit ID, in addition to being higher than 5 bit, was lower than the 3 bit, $t(23) = -5.321$, $p < 0.001$, and the 2 bit, $t(23) = -4.650$, $p < 0.001$. The 3 bit and 2 bit difficulties were similar, $t(23) = 0.724$, $p = 0.470$. In other words, the 3 bit difficulties led to the highest throughput ($M = 2.70$ bits/s), followed by the 2 bit ($M = 2.66$ bits/s), followed by the 4 bit ($M = 2.45$ bits/s), with the 5 bit difficulties having the lowest throughput ($M = 2.15$ bits/s).

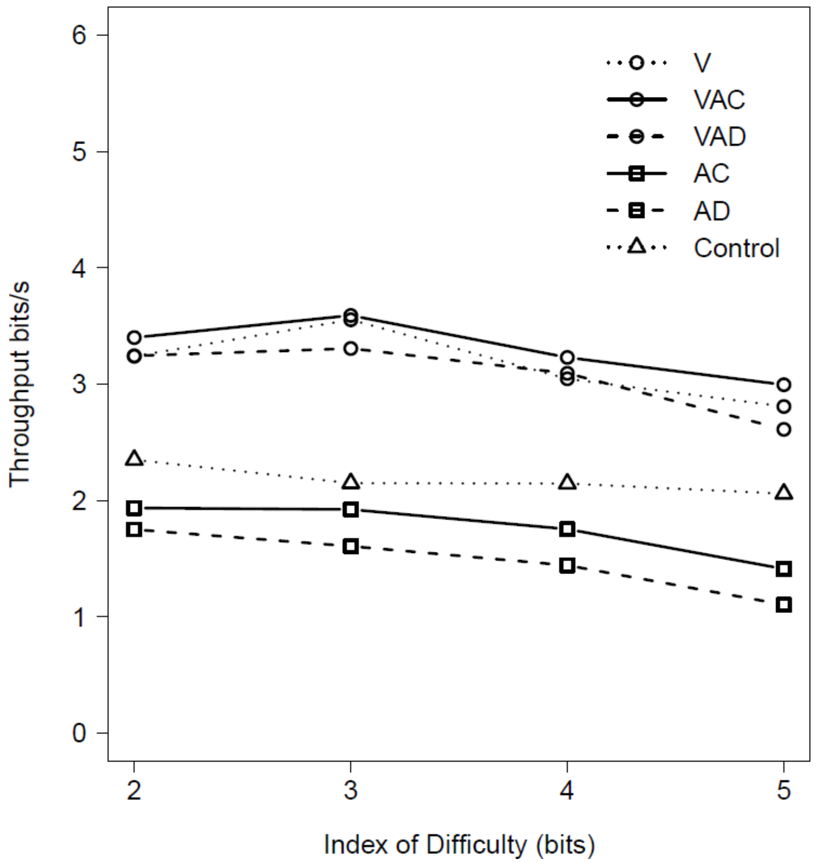

Figure 6.7 average throughput across conditions for each difficulty.

6.5 Discussion

This experiment aimed to uncover the relationship between display modality and target acquisition speed, accuracy, and throughput. Results convey a clear message about the role of visual displays and a more nuanced story about the influence of auditory displays on aimed movements using air gesture controls.

Selection time

As expected, visual displays resulted in faster movement times compared to auditory displays. Previous literature has shown that visual information is readily integrated into trajectory corrections (Elliott, Helsen & Chua, 2001; Zelaznik, Hawkins, & Kisselburgh, 1983), suggesting that using auditory displays to convey information about movement trajectory more effortful, and could lead to slower aimed movements. Since research has demonstrated that the ballistic phase of aimed movements remains relatively unaffected by the availability of visual information, we are forced to conclude that the observed effects of these different conditions can be found in the control phase of movement (Elliot & Hansen, 2001; Elliot et al., 1991, Carson et al., 1993; Chua & Elliot, 1993; Elliot, Lyons & Dyson, 1997), it is likely that the source of the performance differences between the conditions was a result of differences in trajectory corrections made during the control phase of the movements. Elliot and colleagues (1991) showed that the number of trajectory corrections was not influenced by the availability of visual information, implying that the improved accuracy observed with visual displays is a result of more efficient trajectory corrections. Therefore, it is possible that the reason that the audio-only conditions led to slower movements was because the human motor system does not integrate auditory information into online corrections as quickly and easily as with visual information, meaning that the corrective movements made in the control phase of movements are slower. The control condition was also faster than the auditory display conditions because the control condition provided participants only with a visual memory to aim at, meaning that corrective movements in are not possible because there is no new information about the relative position between the hand and target incoming. Interestingly, conditions with visual display all led to similar completion times as the control condition, meaning that visual information can be integrated fast enough that it does not lead to slower movements than the control condition – which requires no corrective movements. Another interesting observation from the study results was that all conditions that had visual information performed similarly, including visual-auditory displays. This suggests that participants were using the visual information to guide their movements while ignoring the auditory displays.

Selection accuracy

Regarding selection accuracy, auditory-only displays led to similar percentages of in-target selections compared to conditions with visual displays, especially at lower levels of difficulty. Interestingly, auditory-only displays consistently resulted in a statistical interaction, showing slower and less accurate movements, especially for much higher difficulty movements (ID = 4, 5) compared to the conditions with visual displays. The control condition led to the lowest accuracy overall, which makes sense given the lack of information provided in that condition. The effect of the difference can be explained by satisficing behavior from participants who found that the amount of time required to make accurate selections for very small targets was not worth it because of the greater effort required to make corrections with auditory-only displays. For larger targets, the information demand was sufficiently low that participants were able to use the limited information from proprioception and/or auditory displays to make accurate selections, but the efficiency of visual displays became more evident for small targets, leading to a slight shift in emphasis in the speed-accuracy tradeoff.

Undershooting and overshooting

Analysis of overshooting and undershooting showed proportionately greater percentage of misses attributed to overshooting for conditions without visual information (AD, AC, Control) compared to conditions with visual information (V, VAD, VAC). Of course, the opposite is also true, that conditions without visual information led to proportionately less undershooting compared to visual conditions. This result suggests that participants undershoot when using visual information. This result is consistent with previous literature that says that participants' initial movements frequently undershoot target positions (Carlton, 1979; Elliot and Chua, 1993). According to the stochastic submovement model (Meyer, Abrams, Kornblum, Wright & Smith, 1988), the distribution of endpoints should vary equally on either side of the target. This difference possibly suggests that visual information is facilitating a speed-emphasized strategy in the speed-accuracy tradeoff. The participants are able to see they are close to the target position and choose to select on the closer side of the target (undershooting) rather than make additional corrections with the possibility of getting closer. This strategy would save participants time, which is consistent with the observation that participants made faster target selections when making selections in conditions with visual displays. In addition, as the follow-up time-after-reaching target analysis confirms, nearly all of the difference in selection time across all conditions was explained by the time participants took after initially reaching the target position. All of this is consistent with the notion that when participants had visual information they were able to choose to make selections without the need for additional corrections requiring them to move further to the right (potential for overshooting). Interestingly, the addition of auditory displays to a visual display led to proportionately more overshooting, more closely resembling the 50/50 overshooting to undershooting ratio expected by the stochastic submovement model (Meyer, Abrams, Kornblum, Wright & Smith, 1988). However, the control condition led to significantly more overshooting compared to all other conditions, with well over 50% overshooting across all difficulty levels. This result suggests that when participants can only rely on a decaying visual memory of target position and kinematic cues, they will overestimate how far away the target position is and will therefore be more likely to overshoot than undershoot. All of these data suggest that participants are able to more easily use a speed-emphasized strategy when making target selections by choosing to miss short (shorter travel distance) rather than long when they have visual information available. The addition of an auditory display appears to cause at least some users to take a little more time to move more to the right, perhaps due to the novelty and saliency of the auditory display, which gave them additional feedback when they were over the target position.

Throughput

Throughput, because it is derived from the accuracy and speed results shows a similar result as the selection time and accuracy results. However, because throughput is highly sensitive to time differences, the control condition actually outperformed the auditory-only displays. Of course, considering the context of this research, we would be wise to more heavily emphasize accuracy because misses while using in-vehicle air gesture systems could lead to frustration and overall greater time completing secondary tasks than selections that take one or two seconds longer.

Comparing between the continuous and discrete auditory-only displays, the continuous auditory display led to faster selection times and comparable accuracy, leading to overall higher throughput, although this difference did not reach statistical significance, it does highlight the potential of a continuous auditory display to outperform a discrete auditory display. This difference, if it is considered a difference at all, is likely a result of continuous auditory displays impacting online trajectory corrections in the control phase of the movement. Existing literature suggests visual information regulates the efficiency of online corrections, and by analogy continuous auditory displays could be serving the same purpose, albeit with less overall efficiency as seen by the relatively small effect size of the difference between discrete and continuous displays. Since both auditory display types led to similar accuracy and continuous auditory displays theoretically could lead to more efficient trajectory corrections that could explain the relatively faster task completion times.

Chapter 7

7 General Discussion

There were a few consistent trends across the first three experiments. First, auditory displays afford drivers the possibility of completing simple tasks without looking away from the road. However, the auditory displays had mixed effects on driving performance, the most important metric when determining the feasibility of an in-vehicle information system design. In the first experiment, the percentage of lane departures was higher than in Experiments 2 and 3. This effect is most likely explained by the change in driving scenario. By adding a lead vehicle, participants had a reference to guide their movement within the lane. However, the lane departures in Experiment 2 were also lower than in Experiment 3. This difference could be explained by a learning effect. Participants in Experiment 2 had only 4 conditions to get through in one hour and had longer periods of driving, allowing more driving time, whereas Experiment 3 had 6 conditions and a longer NASA-TLX so less time was spent driving. But why did the addition of an auditory display not have larger impacts on driving performance? Experiment 3 showed significant improvements in standard deviation of lane position, and both Experiments 1 and 3 showed numerical improvements in lane departure percentage associated with the addition of auditory displays, but why is the effect size so small? One potential explanation is that the visual demands of the driving task could be sufficiently met using peripheral vision because the lane keeping does not require focus visual attention. This would allow drivers to shift their focal visual attention to the secondary task with limited degradation to driving performance.

Another consistent finding was that secondary task performance using air gesture controls was slower for conditions with auditory displays than those with visual displays. This effect was exaggerated for auditory-only displays, which were even slower than auditory-visual prototypes. This highlights a fundamental difference in information capacity between the visual and auditory modalities. However, the fourth experiment hinted at some potential for continuous auditory displays to improve the speed of target acquisitions, although that effect would still not compensate for the time differences seen between audio-only displays and displays with visual information.

A third major trend was that addition of auditory displays led to decreased perceived workload among participants across Experiments 1, 2, and 3. Anecdotal evidence from post-study interviews with participants revealed that the most commonly cited reason for reduced workload was the feeling of comfort being able to keep their eyes on the road when using controls with auditory displays.

7.1 Limitations

Scenario for Experiments 1, 2, and 3

Although the scenario used in Experiments 1, 2, and 3 served their purpose as representative driving environments, future experiments may benefit from adding tasks that require drivers to react to changes in the environment because crash risk is effected more by delayed reaction times associated with visual distraction than degradation in lane keeping performance (Horrey & Wickens, 2006). In future experiments, it would be valuable to include a driving scenario with unexpected events to compare the relative impacts of each prototype on the ability to avoid crashes that result as a function of delayed reaction times, the aspect of driving most degraded by visual distraction.

Control condition for Experiment 4

There is some evidence that suggests that visual memories of target positions can persist for a short period of time after they are removed (Elliot & Jaeger, 1988; Elliot & Madalena, 1987). The control condition in the 4th experiment presented the target positions and then removed them when the participant selected the start button. This means that there is still a trace amount of visual information available to participants as the visual memory decays. It is possible to create a control condition using the same software that includes visual masking to eliminate the possible effect of a trace visual memory of the target position. This could have been achieved by requiring participants to select somewhere within the sensor range after completing a trial. At that point the target could appear. Then after several seconds, the visuals could disappear for a 2 second period and then reveal the start button and no targets. A decay period of 2-10 seconds has been shown to mitigate or eliminate the effect of trace visual memories of target positions (Elliot & Jaeger, 1988; Elliot & Madalena, 1987). This method would ensure that participants did not start the trial until at least two seconds after the visual information was removed. This approach would have to be carried out throughout all of the conditions to remain consistent.

Sound design for Experiment 4

The sound design for conditions that had auditory displays was chosen in order to make a cleaner comparison between a continuous auditory display and a discrete one. Of course, there were many other ways we could have chosen to design the auditory displays. However, since the goal was to compare a discrete and a continuous display with the ultimate goal of informing better in-vehicle auditory display design. We deliberately avoided creating a tempo-based proximity sonification that beeped faster as the user's hand approached the target because we thought it would be too annoying to drivers to hear as part of a secondary task. The goal of the sound design we used was not to create sounds that were aesthetically pleasing, but rather represent a sound design archetype – something that could be made more appealing without undermining the transfer of information. It was our assumption that even an aesthetically pleasing version of a tempo-based sonification would be too distracting for drivers. Researchers have previously suggested amplitude (loudness) and low-pass filters can be used to effectively convey distance to a target area in a large complex auditory display (Gaver & Smith, 1990). In this example, Gaver and Smith used a negative polarity, meaning that the sounds get louder as the distance to the target gets smaller. This metaphor reflects most of our daily experience in which sounds get louder as we get closer to their source. For these reasons, it is a good candidate for consideration in future sound designs related to this project. Walker (2002; 2007) compared the appropriateness of tempo, frequency, and modulation parameter mappings for magnitude estimation of different types of data, including temperature, velocity, pressure, size, and *proximity* – the key metric for sound design in

Experiment 4. Walker (2007) tested proximity estimation using tempo, frequency, and modulation in two identical experiments, recruiting hundreds of participants, all from the same population. In the first experiment, frequency outperformed both tempo and modulation. In the second experiment, an exact replication of the first experiment, recruiting from the same subject pool, both tempo and modulation index outperformed frequency. In addition, the performance of positive and negative polarities was inconsistent. In Walker's first experiment, for frequency mappings, a positive polarity outperformed a negative polarity, but the opposite was true in the replication (2007). All of this points to the fact that the effectiveness of sonification is highly variable and depends on the situation and the sound design. Therefore, it is possible that a tempo-based mapping could have resulted in better performance. It is very hard to say with any certainty what the best design would be based on past experimental results.

Also, since the goal was to create a fully orthogonal experimental setup, comparing visual-only, auditory-only, and visual plus auditory displays for each type of sound design, it would be highly impractical to evaluate tempo-based, and modulation-based mappings as well. In the future, it would be helpful to make comparisons between these different types of continuous displays.

Models of motor control

When there is no visual information, or if the movements are very fast (too fast for visual processing to influence movement), the impulse variability model of motor control provides a good explanation for endpoint determination (Wallace & Newell, 1983). The results of Experiment 4 suggest that auditory information can be used to guide movements. Wallace and Newell's statement (1983) that non-visual movements can be explained well by the impulse variability model is undermined by the relatively high accuracy and slower selection times found in the auditory-only display conditions compared to the control condition (no information). Wallace and Newell's statement arose from their observations that movements made with no information are relatively fast (too fast for visual processing). This suggests that participants were able to use auditory information to guide their movements which is why their movements were slower and more accurate. Instead of relying solely on the initial impulse, participants appear to have been making secondary corrections based on the auditory information. However, it also appears that auditory information is processed more slowly compared to the visual information because selection times for the auditory-only conditions were slower than conditions with visual information. This difference between the visual and auditory-only conditions also appears exaggerated at higher difficulty levels. This result is consistent with the observations of previous research that says the importance of visual information is greater for movements that require high accuracy (higher index of difficulty).

One possible explanation for the relatively lower accuracy observed for the auditory-only conditions is that auditory information requires longer to process than visual information. Elliot et al. (1991) suggested that the extra time was a result of visual processing, done with the purpose of reducing target-aiming error. Previous literature has shown that tactile information is processed more quickly than auditory information in target acquisition tasks without degrading error rates (Akamastu, MacKenzie, & Hasbrouq, 1995; Sharmin, Evreinov, & Raisamo, 2005; Charoenchaimonkon, Janecek, Dailey, & Suchato, 2010). Charoenchaimonkon and colleagues (2010) showed that visual feedback led to faster selection times and reduced error rates compared to auditory or tactile displays. Akamastu et al. (1995) found no difference between auditory and visual feedback for selection times. It is noteworthy that the focus of all of these studies was on use of feedback after users position the cursor in the target, rather than in-the-loop information that can be used to guide small trajectory corrections in real-time over the course of the movement. However, there seems to be some indication that visual processing could facilitate faster target selections compared to auditory-only feedback, at least in some circumstances.

Experiment 4 Apparatus

The LEAP Motion sensor had the capability of recording hand position data at approximately 100 Hz. However, the time resolution was variable, meaning sometimes the gap between data points could be $1/100^{th}$ of a second, other times could be $8/100^{th}$ of a second. This inconsistency led to temporal lag issues which could potentially explain some of the difficulty participants had using the air gesture systems, i.e., why the selection accuracies were generally in the low to mid ninety percent range rather than near one hundred percent as was observed for the touchscreens.

Generalizability

It is difficult to generalize very far beyond the context of these studies due to the limits of the prototype designs used during the study. Different effects may emerge in conditions where menu items are not already memorized or contain many more items, such as long lists of songs.

7.2 Future Work

There are a number of interesting questions that emerged from this line of research, so many that they could not be covered in this dissertation. I will discuss each of those questions, as they pertain to in-vehicle gesture controls. First, Experiment 1 highlighted a need to investigate the range of space in which users can easily complete goal-directed acquisitions. My research already showed some performance degradation associated with movements close to the body, which does not conform to predictions from Fitts's Law, which says that closer movements, with equivalent target sizes, should be faster and more accurate. If I knew how goal-directed movement performance is influenced by the location in space around the user, I could better know where to place sensors and define ranges in which controls should be placed for optimal human performance.

Many researchers have demonstrated that maximum reach envelopes can be described as spherical surfaces (e.g., Figure 7.1; Kennedy, 1964; Chaffee, 1969; Klein, 2012). This has been taken into account in the automotive standard SAE J287, which presents population reach envelopes to ensure that in-vehicle controls are not designed out of reach of drivers within 5th to 95th percentiles (1988; 2007). However, the SAE J287 (1988) reach envelope standard is not helpful for the design of a gesture control system because it describes reach envelopes on only two dimensions, and does not include depth because it assumes all controls are positioned on the front dashboard of the vehicle.

Recently, researchers have gone further in their investigation of seated reach envelopes to consider the difficulty when defining seated reach limits (Yu et al., 2017; Liu et al., 2017). They concluded that the SAE J287 reach envelopes are not consistent with comfort envelopes, meaning that some of the positions of controls allowed by the SAE J287 standard are uncomfortable for some drivers. Yu et al. (2017) also modeled the reach envelopes associated with a continuum of reaching difficulties and found that the J287 standard guidelines resulted in reach envelopes that were at a 5 or 6 on an 11 point scale. Thus, even if the gesture controls had passed the J287 standard expectations, target acquisitions may have been difficult or uncomfortable.

As can be seen in Figure 7.1, there appears to be a pocket of space to the immediate right of the pilot on the horizontal plane, within about 40 cm, where there is a much wider variance in reach envelope between individuals (Figure 7.1, left). From a vertical cross section, looking from the back of the pilot, again, there appears to be a small pocket where there is more highly variable reaching capability within about 40 cm to the right of the pilot from about shoulder height to seat height (Figure 27, right). Kennedy's research (1964) demonstrated that there is a lot of variability in people's ability to reach some areas that are close to the body. This result offers an explanation for the unexpectedly poor target selection performance for targets on the very bottom left in Experiment 1. For those targets, because the controls were horizontal for all conditions, targets on the bottom left would be closer to the participant's body. Since the position of the sensor remained fixed (to keep a constant distance from the steering wheel), some participants may have found the position required to select those targets to be difficult or impossible to reach. This result would align with the high variability in reach envelopes seen in Kennedy's research.

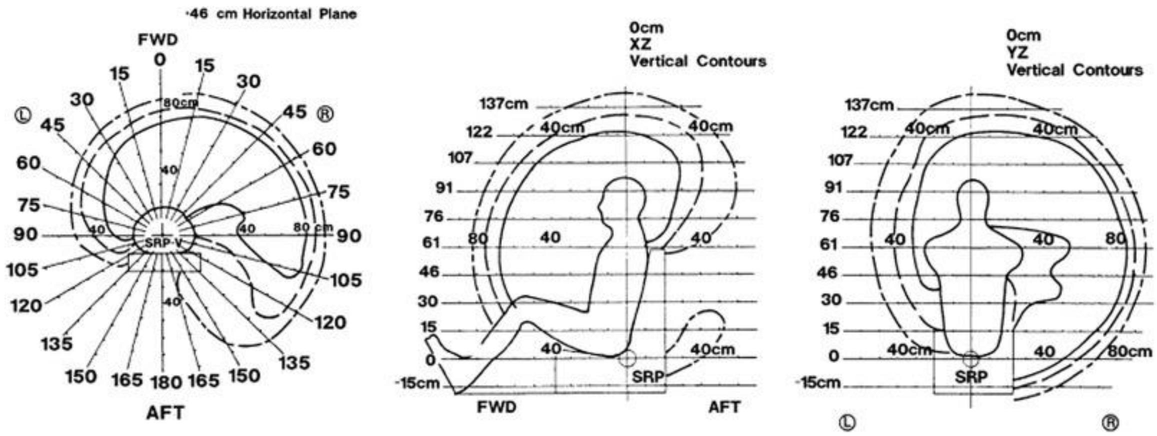

Figure 7.1 reach envelopes measured by Kennedy (1964) showing 5th, 50th, and 90th percentiles of maximum reach. [permission not required as it is public domain]

Control condition

One obvious gap in this research is the lack of a control condition in Experiments 1, 2, and 3. While the inclusion of a control condition would allow for comparisons to baseline driving performance (without any secondary tasks, it would also not help improve the design of in-vehicle gesture controls. Driving performance will often be degraded by the addition of any non-driving-related task, including those required to operate in-vehicle information systems available in nearly all production vehicles (e.g., radio knobs, seat controls, mirror adjusters). It appears that the benefits of certain technological additions to vehicles outweigh the potential degradation to driving performance. Assuming that drivers, manufacturers, and regulators are collectively willing to accept the risks associated with adding these technologies to vehicles, the real benchmark is not to compare to driving without distraction, but rather to compare to distractions we have already accepted, e.g., radio knobs, etc. So, the touchscreen comparison made in Experiment 2 serves as a surrogate control condition because touchscreen interfaces are widely used in all vehicles with acceptance from regulatory bodies and the general population.

Kinematic features

Research has shown that when visual information is provided, participants spend more time in the corrective phase of movement (Elliot, Garson, Goodman & Chua, 1991; Carson, Goodman, Chua & Elliot, 1993; Elliot, Lyons & Dyson, 1997). It is not clear from these experiments if the same patterns hold true for auditory-only displays. It could be a subject of future research to examine more closely the kinematic features associated with visual and auditory-guided target acquisitions including skewness of velocity profiles and number of trajectory corrections. This line of research might be of interest to researchers interested in motor control models for how information guides movements in goal-direction aiming tasks.

In order to answer research questions around the kinematic features of target acquisitions use of a higher-fidelity, more precise kinematic sensor is required. The LEAP Motion suffers from both spatial and temporal resolution limitations in comparison to much more expensive systems like a Vicon tracking system. Using a system like the Vicon tracking system may provide the opportunity to gain more insight into time-series data and selection accuracies for smaller targets.

Real world use

Another key issue is measuring the real-world performance of the current prototype systems. I have two major questions in that regard: (1) Our participants used the system over only a 30-40 minute period for each of our experiments. How much better could they perform if they had several hours to practice, or if they used it over an extended period of time? What does the learning curve look like? These are unanswered questions that would be a good topic for future research. (2) Another remaining question is, what does real world use of a gesture control system look like? The driving simulator and laboratory conditions control for many of the realities of day-to-day driving. In the real world, drivers are putting on makeup, eating food, or checking their phones while driving. How could these other non-driving activities impact the use of a gesture control system? This is another good topic for future research.

7.3 Design Guidelines

This section contains several general guidelines derived from things learned through the four experiments completed as part of this dissertation.

1. Provide auditory/visual displays in combination.

   Although much of this research is directed at the use of auditory displays as a means of communicating information in vehicles, the flexibility afforded by combinations of auditory and visual displays allows for users to adopt an eyes-free approach as they can learn to trust the system (if ever).

2. Provide equivalent auditory information for every piece of visual information to all users to use eyes-free.

If the goal is to replace dependency on visual information, auditory information should mirror the visual information as much as possible, so users have the opportunity to search through menu without missing any information they may rely on to navigate through the menu.

3. Consider using a continuous auditory display to support users who do not trust discrete auditory displays.

   Use of a continuous auditory display can provide fine-grain information about the relative position of cursor and target center/edge to give users an added layer of information, ultimately allowing users to get as much information from the auditory display as possible. If taking this route, special care should be taken to minimize auditory clutter that can come with using a continuous auditory display.

4. If using an auditory-only display, consider the additional time cost that comes with higher difficulty movements (smaller/more distant targets) compared to visual displays.

   Target selections always take longer at higher difficulties than at lower difficulty levels. That relationship is known as Fitt’s law (Fitts, 1954). However, the slope of that line is steeper, with auditory-only displays leading to relatively slower selections at higher difficulties compared to visual-only and visual-auditory displays. This extra time is of greater concern for in-vehicle menus with more depth, meaning more selections will be made.

5. Keep the ID under 2.5 to ensure highest possible accuracy rates.

As observed in the first experiment, and confirmed in experiment 2 and 3, the 2x2 grids led to accuracy rates around 90% (ID less than 2.5 bits) while the 4x4 grids led to accuracy rates around 65% (ID greater than 2.5 bits).

6. Consider uneven target acquisition performance within reach envelope in vehicle cabin. Closer targets are not always resulting in the best performance.
   As was observed in Experiment 1, there were pockets of space in the air that participants had difficulty reaching and that degraded secondary task accuracy rates. The space close to the right hip of participants led to degraded target selections. This may not be the only position in space that leads to degraded target acquisition performance. However, consideration needs to be made even for this one difficult position alone because it is one of the areas that could be a candidate for positioning a sensor in a vehicle cabin.

The NHTSA guidelines suggest that the visual demands to complete secondary tasks with in-vehicle menus should be able to be completed with glances shorter than 2 seconds (National Highway Safety and Traffic Administration, 2012). This guideline was met for each of the experiments. Only a small minority of glances were greater than 2 seconds. The other guideline from NHTSA stated that drivers should not exceed the frequency of lane departures in comparison to a reference task (e.g., tuning a radio). The second experiment demonstrated that air gesture controls led to more lane departures (more time-out-of-lane) compared to touchscreen use, which can be considered a commonly accepted in-vehicle control that could serve as a reference task. While this standard suggests that air gesture controls are in violation of a basic guideline, there is still a lot of potential in air gesture controls that remains untapped by the designs used in these studies. For example, the auditory-only display may have the greatest potential to alleviate visual demands and it also showed the greatest ability to reduce lane departures relative to a visual-only display. However, that system was not tested against an equivalent touchscreen system. As an aside, there could be some methodological issues with the guideline requirements set by NHTSA because some new in-vehicle technologies may be less familiar to study participants and results may be degraded for unfamiliar systems due to unfamiliarity rather than the impact of the technology use on driving after reaching peak performance. Another point of contention relates to the performance metric. The frequency of lane departures could potentially be influenced by the demands of the driving scenario. If for example, drivers completed a driving task that only required them to drive straight the visual demands may be quite low and sufficiently met with use of ambient visual attention. On the other hand, if the driving scenario had curvy roads, such as in these experiments, the visual demand may be greater. Many other road conditions, such as the presence of lane lines, other vehicles, turning decisions, etc. might all influence the visual attention requirements to maintain good lane keeping. The point here is that it could be possible to design a driving scenario in such a way that driving performance could reach a ceiling effect and performance would remain equivalent for in-vehicle technologies that actually require very different levels of visual attention to use. It is possible that, had Experiment 2 been used a straight road, driving performance would have been equivalent between the two, and the air gesture control system would have met the performance guideline requirement. Perhaps, the NHTSA (2012) in-vehicle control design guidelines should include additional methodological requirements or recommendations for testing the relative lane deviation performance in order to limit these potential biases.

7.4 Conclusion

In Experiment 1, I found that larger target sizes and smaller number of targets in the 2x2 grid, resulted in improved performance in lane control, secondary task performance, reduced the number of eye glances away from the road, and reduced driver workload compared to the 4x4 grid. I found that for menus arranged in a square grid, 16 smaller targets (IDs between 2.43-3.07) were more difficult to select than 4 larger targets (IDs 1.77-2.13) while driving which led to difficulties in multitasking. The addition of an auditory display reduced the frequency of off-road glances – lowest for 2x2 grids, but greater reduction was observed in 4x4 grids – and lowered driver workload, especially for 2x2 grids, but some participants found auditory feedback annoying for 4x4 grids. In addition, I found that the position of targets in 4x4 grids resulted in unexpectedly slow and inaccurate selection times for targets in the closest corner of the grid, highlighting a need to measure in-air target acquisition performance within the reach envelope of the driver.

Experiment 2 demonstrated that the 2x2 auditory-supported air gesture systems resulted in fewer off-road glances, and resulted in comparable perceived workload but did result in relatively more time spent out-of-lane compared to touchscreen controls. Results also showed participants generally preferred auditory-supported gesture controls over touchscreen controls. Deeper analysis suggested that some participants looked away from the road a lot more than others, highlighting a potential vein of future research investigating individual differences in user acceptance of auditory-supported gesture controls.

Experiment 3 demonstrated that auditory-only displays are feasible, and can be used safely while driving, even showing some incremental improvements over visual-only displays in driving performance. We also observed a tradeoff between eyes-on-road time and secondary task completion times, with auditory-only displays leading to more eyes-on-road time but slower selection times and vice versa for visual/auditory and visual-only displays. Future work will focus on more realistic driving scenarios and more complex secondary tasks to further investigate the potential of auditory-supported air gesture controls in vehicles.

Results from Experiment 4 indicate that auditory-only displays are not as effective as visual displays at guiding aimed movements in target acquisition tasks among sighted users. However, the data suggest that targets can be selected with similar levels of accuracy when using auditory-only displays, especially when movements are less difficult (ID = 2, 3). This suggests the potential for using auditory displays (continuous or discrete) for facilitating eyes-free target acquisitions using air gesture controls. For example, in vehicle contexts, auditory-only displays can result in the same accurate performance in the secondary gesture task, while maintaining visual attention on the road. Therefore, further applied research is required to identify the relationship among the task demand (e.g., level of difficulty), multi-modalities, and different types of auditory displays.

References

Aiken, E. G., & Lau, A. W. (1966). Memory for the pitch of a tone. Perception & Psychophysics, 1, 231–233.

Aiken, E. G., Shennum, W. A., & Thomas, G. S. (1974). Memory processes in the identification of pitch. Perception & Psychophysics, 15, 449–452.

Akamastu, M., MacKenzie, I. S., & Hasbrouq, T. (1995). A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics, 38,* 816-827.

Akyol, S., Canzler, U., Bengler, K., & Hahn, W. (2000, November). Gesture Control for Use in Automobiles. In *MVA* (pp. 349-352).

Alm, H., & Nilsson, L. (1995). The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident Analysis & Prevention,27*(5), 707-715.

Alpern, M., & Minardo, K. (2003, April). Developing a car gesture interface for use as a secondary task. In CHI'03 extended abstracts on Human factors in computing systems (pp. 932-933). ACM.

Balakrishnan, R. (2004). "Beating" Fitts' law: virtual enhancements for pointing facilitation. *International Journal of Human-Computer Studies*, *61*(6), 857-874.

Ball, K., V.G. Wadley, and J.D. Edwards (2002). Advances in technology used to assess and retrain older drivers. Gerontechnology, 2002. 1(4): p. 251-261.

Beilharz K. Gesture-controlled interaction with aesthetic information sonification. InProceedings of the second Australasian conference on Interactive entertainment 2005 Nov 23 (pp. 11-18). Creativity & Cognition Studios Press.

Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. Perception & Psychophysics, 60, 1357–1373.

Bonebright, T. L., Nees, M. A., Connerley, T. T., & McCain, G. R. (2001). Testing the effectiveness of sonified graphs for education: A programmatic research project. In Proceedings of the International Conference on Auditory Display (ICAD01) (pp. 62–66). Espoo, Finland: Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology.

Bonebright, T. L., & Nees, M. A. (2007). Memory for auditory icons and earcons with localization cues. In Proceedings of the Thirteenth Meeting of the International Conference on Auditory Display (ICAD 2007) (pp. 419–422). Montreal, Canada: McGill University, Schulich School of Music.

Bootsma, R. J., Fernandez, L., & Mottet, D. (2004). Behind Fitts' law: kinematic patterns in goal-directed movements. *International Journal of Human-Computer Studies*, *61*(6), 811-821.

Boyle, L. N., Lee, J. D., Peng, Y., Ghazizadeh, M., Wu, Y., Miller, E., & Jenness, J. (2013). Text reading and text input assessment in support of the NHTSA visual-manual driver distraction guidelines (No. DOT HS 811 820).

Boyle, L. N., Lee, J. D., Peng, Y., Ghazizadeh, M., Wu, Y., Miller, E., & Jenness, J. (2013). *Text reading and text input assessment in support of the NHTSA visual-manual driver distraction guidelines* (No. DOT HS 811 820).

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press

Brewster, S. (1998a). The design of sonically-enhanced widgets. *Interacting with computers*, *11*(2), 211-235.

Brewster, S. (1998b). Sonically-Enhanced Drag and Drop. ICAD 98, pp. 1-7.

Brogmus, G. E. (1991, September). Effects of age and sex on speed and accuracy of hand movements: And the refinements they suggest for Fitts' Law. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 35, No. 3, pp. 208-212). SAGE Publications.

Burke, J., Prewett, M., Gray, A., Yang, L., Frederick, S., Coovert, M., Elliot, L., and Redden, E. (2006). Comparing the effects of visual-auditory and visual-tactile feedback on user performance a meta analysis. *Proceedings of the 8th international conference on multimodal interfaces*, p. 108-117.

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.

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.

Cairnie, I. W. Ricketts, S. J. Mckenna And G. Mcallister, A. (2000). Using finger-pointing to operate secondary controls in automobiles. Intelligent Vehicles Symposium, 550-555.

Carlton, L. G. (1979). Control processes in the production of discrete aiming responses. *Journal of Human Movement Studies*, *5*(3), 115-124.

Carlton, L. G. (1981). Visual information: The control of aiming movements.*The Quarterly Journal of Experimental Psychology*, *33*(1), 87-93.

Carson, R. G., Goodman, D., Chua, R., & Elliott, D. (1993). Asymmetries in the regulation of visually guided aiming. *Journal of motor behavior*, *25*(1), 21-32.

Chaffee, J.W., 1969. Methods for Determining Driver Reach Capability. Society of Automotive Engineers. Technical Paper No. 690105. Society of Automotive Engineers Inc.

Charoenchaimonkon, E., Janacek, P., Dailey, M., Suchato, A. (2010). A comparison of audio and tactile displays for non-visual target selection tasks. International Conference on User Science and Engineering (i-USEr).

Chua, R., & Elliott, D. (1993). Visual regulation of manual aiming. Human Movement Science, 12(4), 365-401.

Cockburn, A. and Brewster, S., (2005). Multi Modal Feedback for the Acquisition of Small Targets. *Ergonomics*, *48*(9), 1129-1150.

Cockburn, A., & Gutwin, C. (2009). A predictive model of human performance with scrolling and hierarchical lists. *Human-computer interaction*, 24, 273-314.

Cockburn, A., Quinn, P., Gutwin, C., Ramos, G., & Looser, J. (2011). Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback. *International Journal of Human-Computer Studies*, *69*(6), 401-414.

Crossman, E. R. F. W., & Goodeve, P. J. (1983). Feedback control of hand-movement and Fitts' law. *The Quarterly Journal of Experimental Psychology*,*35*(2), 251-278.

de Grosbois, John, Matthew Heath, and Luc Tremblay. "Augmented feedback influences upper limb reaching movement times but does not explain violations of Fitts' Law." Frontiers in psychology 6 (2015).

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J. D., ... & Bucher, C. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment* (No. HS-810 593).

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.

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 (2010), 403–409

Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

Driver Focus-Telematics Working Group. (2002). Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems. Washington, DC: Automotive Alliance

Drury, C. G. (1975). Application of Fitts' Law to foot-pedal design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *17*(4), 368-373.

Edwards, A. D. (1989). Soundtrack: An auditory interface for blind users.*Human-Computer Interaction*, *4*(1), 45-66.Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of experimental psychology,47(6), 381.

Edworthy, J. (1998). Does sound help us to work better with machines? A commentary on Rautenberg's paper "About the Importance of Auditory Alarms During the Operation of a Plant Simulator." Interacting With Computers, 10, 401–409.

Effenberg, A. (2001). Multimodal convergent information enhances perception accuracy of human movement patterns. In *in Proc. 6th Ann. Congress of the ECSS, Sport und Buch, Strauss*.

Effenberg, A. O., & Mechling, H. (2003). Multimodal convergent information enhances reproduction accuracy of sport movements. In *Proceedings of the 8th Annual Congress of the European College of Sport Science. Salzburg, Austria* (p. 196).

Elliott, D., & Hansen, S. (2010). Visual regulation of manual aiming: a comparison of methods. *Behavior research methods, 42*(4), 1087-1095.

Elliott, D., & Jaeger, M. (1988). Practice and the visual control of manual aiming movements. *Journal of Human Movement Studies, 14*(6), 279-291.

Elliott, D., & Madalena, J. (1987). The influence of premovement visual information on manual aiming. *The Quarterly Journal of Experimental Psychology*, *39*(3), 541-559.

Elliott, D., Garson, R. G., Goodman, D., & Chua, R. (1991). Discrete vs. continuous visual control of manual aiming. *Human Movement Science*,*10*(4), 393-418.

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.

Elliott, D., Lyons, J., & Dyson, K. (1997). Rescaling an acquired discrete aiming movement: Specific or general motor learning?. *Human Movement Science*, *16*(1), 81-96.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of experimental psychology,47(6), 381.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. Journal of experimental psychology, 67(2), 103.

Gable, T. M., Raja, S. R., Samuels, D. P., & Walker, B. N. (2015, September). 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 (pp. 297-304). ACM.

Gaver, W. W. (1989). The SonicFinder: An interface that uses auditory icons.*Human-Computer Interaction*, *4*(1), 67-94.Green, P. (1999). *Visual and task demands of driver information systems* (No. UMTRI-98-16,).

Gelfand, S. A. (2009). Essentials of audiology. New York, NY: Thieme Medical.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002, March). Eye tracking in web search tasks: design implications. In*Proceedings of the 2002 symposium on Eye tracking research & applications* (pp. 51-58). ACM.

Granka, L. A., Joachims, T., & Gay, G. (2004, July). Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 478-479). ACM.

Green, P. (2000, November). Crashes induced by driver information systems and what can be done to reduce them. In Sae Conference Proceedings (pp. 27-36). SAE; 1999.

Green, P. (2004). Driver distraction, telematics design, and workload managers: Safety issues and solutions.

Grossman, T., & Balakrishnan, R. (2004, April). Pointing at trivariate targets in 3D environments. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 447-454). ACM.

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.

Hart, S. G., (2006). NASA Task Load Index (NASA-TLX); 20 years later. Proceedings of Human Factors and Ergonomics Society, 50(9), p. 904-908.

Hatfield, B. C., Wyatt, W. R., & Shea, J. B. (2010). Effects of auditory feedback on movement time in a Fitts task. Journal of motor behavior, 42(5), 289-293.

Heath, M., Hodges, N. J., Chua, R., & Elliott, D. (1998). On-line control of rapid aiming movements: Unexpected target perturbations and movement kinematics. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *52*(4), 163.

Hick, W. E., (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11-26.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & psychophysics*, *57*(6), 787-795.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(1), 196-205.

Horrey, W., & Wickens, C. (2007). In-vehicle glance duration: distributions, tails, and model of crash risk. Transportation Research Record: Journal of the Transportation Research Board, (2018), 22-28.

Hurwitz, J. B., & Wheatley, D. J. (2002, September). Using driver performance measures to estimate workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, No. 22, pp. 1804-1808). SAGE Publications.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188-196.

Jacko, J., Salvendy, G. and Koubek, R. (1996). Modeling of Menu Design in Computerized Work. *Interacting with Computers*, 7(3), 304-330.

Jeon, J. Y., & Fricke, F. R. (1997). Duration of perceived and performed sounds. Psychology of Music, 25, 70–83

Jeon, M., & Walker, B. N. (2009, October). “Spindex”: Accelerated Initial Speech Sounds Improve Navigation Performance in Auditory Menus. InProceedings of the Human Factors and Ergonomics Society Annual Meeting(Vol. 53, No. 17, pp. 1081-1085). SAGE Publications.

Jeon, M., & Walker, B. N. (2011). Spindex (speech index) improves auditory menu acceptance and navigation performance. ACM Transactions on Accessible Computing (TACCESS), 3(3), 10.

Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., 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.

Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological bulletin*, *70*(6p1), 387.

Keele, S. W., & Posner, M. I. (1968). Processing of visual feedback in rapid movements. *Journal of experimental psychology*, *77*(1), 155.

Kennedy, K.W., 1964. Reach Capability of the USAF Population, Phase I, the Outer Boundaries of Grasping-reach Envelopes for the Shift-sleeved, Seated Operator. Technical Report AMRL-TDR-64-59, Ohio. Aerospace Medical Research Laboratories Wright-Patterson Air Force Base.

Kerr, R. (1973). Movement time in an underwater environment. *Journal of Motor Behavior*, *5*(3), 175-178.

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.

Klein, B.A., 2012. Development of a Statistical Model of Reach Exertion Magnitude Perception for Use in Designing Truck Cabs. Master Thesis. Pennsylvania State University. https://etda.libraries.psu.edu/catalog/14497.

Kramer, G. (1993). *Auditory display: Sonification, audification, and auditory interfaces*. Perseus Publishing.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), Auditory display: Sonification, audification, and auditory interfaces (pp. 1–78). Reading, MA: Addison-Wesley

Kramer, G., Walker, B. N., Bonebright, T. L., Cook, P., Flowers, J., Miner, N., Neuhoff, J., Bargar, R., Barrass, S., Berger, J., Evreinov, G., Fitch, W. T., Gröhn, M., Handel, S., Kaper, H., Levkowitz, H., Lodha, S., ShinnCunningham, B., Simoni, M., & Tipei, S. (1999). The sonification report: Status of the field and research agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. Santa Fe, NM: International Community for Auditory Display

Langhoff, P. W., Corcoran, C. T., Sims, J. S., Weinhold, F., & Glover, R. M. (1976). Moment-theory investigations of photoabsorption and dispersion profiles in atoms and ions. *Physical Review A*, *14*(3), 1042.

Lank, E., Cheng, Y. C. N., & Ruiz, J. (2007, April). Endpoint prediction using motion kinematics. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 637-646). ACM.

Lansdown, T. C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. *Ergonomics*, *47*(1), 91-104.

Liu, Y.-C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. Ergonomics, 44, 425–442.

Liu, Q., Ren, J. Qian, Z., Hua, M., (2017). Seated reach capabilities for ergonomic design and evlaution with consideration of reach difficulties. *Applied Ergonomics*, 59(A), 357-363.

MacKenzie, C. L., Marteniuk, R. G., Dugas, C., Liske, D., & Eickmeier, B. (1987). Three-dimensional movement trajectories in Fitts' task: Implications for control. *The Quarterly Journal of Experimental Psychology*, *39*(4), 629-647.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. Human-computer interaction, 7(1), 91-139.

Manes, D., Green, P. (1997). Evaluation of a Driver Interface: Effects of Control Type (Knob Versus Buttons) and Menu Structure (Depth Versus Breadth). University of Michigan Transportation Research Institute (UMTRI): Technical Report UMTRI-97-42.

Marentakis, Georgios, and Stephen A. Brewster. "A study on gestural interaction with a 3d audio display." International Conference on Mobile Human-Computer Interaction. Springer Berlin Heidelberg, 2004.

May, K. R., Gable, T. M., & Walker, B. N. (2014, September). 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* (pp. 1-6). ACM.

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.

McKeown, D., & Isherwood, S. (2007). Mapping candidate within-vehicle auditory displays to their referents. Human Factors, 49, 417–428.

Medina, J., Jax, S. A., & Coslett, H. B. (2009). Two-component models of reaching: Evidence from deafferentation in a Fitts' law task. Neuroscience letters, 451(3), 222-226.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Keith Smith, J. E. (1988). Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological review*, *95*(3), 340.

Miller, D. (1981). The Depth/Breadth Tradeoff in Hierarchical Computer Menus. Proceedings of the Human Factors Society25th Annual Meeting, 25(1), 296-300.

Mottet, D., & Bootsma, R. J. (1999). The dynamics of goal-directed rhythmical aiming. *Biological cybernetics*, *80*(4), 235-245.

Mynatt, E. D., & Edwards, W. K. (1992). The Mercator environment: A nonvisual interface to X Windows and Unix workstations. Puckette, M. S. (1996) Pure Data. Proceedings, International Computer Music Conference. San Francisco: International Computer Music Association, pp. 269–272.

Myers, R., Ball, K., Kalina, T., Roth, D., & Goode, K. (2000). Relation of useful field of view and other screening tests to on-road driving performance. *Perceptual and Motor Skills*, 91, 279-290.

National Highway Traffic Safety Administration. (2012). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices.*Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT)*.

Nees, M., Walker, B. (2011). Auditory Displays for In-Vehicle Technologies. *Reviews of Human Factors and Ergonomics, 7*(1), 58-99.

Nielsen, J., & Pernice, K. (2010). *Eyetracking web usability*. New Riders.

Ohn-Bar, E., Tran, C., & Trivedi, M. (2012, October). Hand gesture-based visual user interface for infotainment. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 111-115). ACM.

Olson, R. L., Hanowski, R. J., Hickman, J. S., & Bocanegra, J. L. (2009). Driver distraction in commercial vehicle operations (No. FMCSA-RRR-09-042).

Peng, Y., Boyle, L., and Hallmark, S., (2013). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. *In Accident Analysis and Prevention*.

Pierno, A. C., Caria, A., & Castiello, U. (2004). Comparing effects of 2-D and 3-D visual cues during aurally aided target acquisition. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(4), 728-737.

Pierno, A. C., Caria, A., Glover, S., & Castiello, U. (2005). Effects of increasing visual load on aurally and visually guided target acquisition in a virtual environment. *Applied ergonomics, 36*(3), 335-343.

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. MPH, 69–75.

Ranney, T. A., Mazzae, E., Garrott, R., & Goodman, M. J. (2000, July). NHTSA driver distraction research: Past, present, and future. In *Driver distraction internet forum* (Vol. 2000).

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine*, *336*(7), 453-458.

Riener, A. (2012). Gestural interaction in vehicular applications. Computer,45(4), 42-47.

Rosenbaum, D. A. (2009). *Human motor control*. Academic press.

Society of Automotive Engineers, 1988. Recommended Practice J287, Driver Hand Control Reach. Society of Automotive Engineers, Inc, Warrendale PA.

Society of Automotive Engineers, 2007. Recommended Practice J287, Driver Hand Control Reach. Society of Automotive Engineers, Inc, Warrendale PA.

Schmidt, R. A., Zelaznik, H. N., & Frank, J. S. (1978). Sources of inaccuracy in rapid movement. *Information processing in motor control and learning*, 183-203.

Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S., & Quinn Jr, J. T. (1979). Motor-output variability: a theory for the accuracy of rapid motor acts.*Psychological review*, *86*(5), 415.

Shakeri, G., Williamson, J. H. and Brewster, S. (2017) Novel Multimodal Feedback Techniques for In-Car Mid-Air Gesture Interaction. In: AutomotiveUI 2017, Oldenburg, Germany, 24-27 Sep 2017, pp. 84-93.

Sharmin, S., Evreinov, G. and Raisamo, R. (2005). Non-Visual Feedback Cues for Pen Computing. Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.

Snowberry, Parkinson, and Sisson (1983). Computer Display Menus. *Ergonomics*, 26, 699-712.

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.

Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI.*International journal of human-computer studies*, *61*(6), 751-789.

Srinivasan, R., & Jovanis, P. P. (1997). Effect of selected in-vehicle route guidance systems on driver reaction times. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 200-215.

Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: Loudness. Psychological Review, 43, 405–416.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. Journal of the Acoustical Society of America, 8, 185–190.

Strayer, D. L., & Drew, F. A. (2004). Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(4), 640-649.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone.*Psychological science*, *12*(6), 462-466.

Stutts, J. C., Reinfurt, D. W., Staplin, L., & Rodgman, E. A. (2001). The role of driver distraction in traffic crashes.

Sugden, D. A. (1980). Movement speed in children. *Journal of Motor Behavior*, *12*(2), 125-132.

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

Tsimhoni, O., & Green, P. (2001, October). Visual demand of driving and the execution of display-intensive in-vehicle tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 45, No. 23, pp. 1586-1590). SAGE Publications.

Tsimhoni, O., Smith, D., & Green, P. (2004). Address entry while driving: Speech recognition versus a touch-screen keyboard. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(4), 600-610.

Wade, M. G., Newell, K. M., & Wallace, S. A. (1978). Decision time and movement time as a function of response complexity in retarded persons.*American Journal of Mental Deficiency*.

Walker, B., (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, *8*(4), 211-221.

Walker, B. N., & Lindsay, J. (2006). The effect of a speech discrimination task on navigation in a virtual environment. In Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting (pp. 1538–1541). Santa Monica, CA: Human Factors and Ergonomics Society.

Walker, B., (2007). Consistency of magnitude estimations with conceptual data dimensions used for sonification. *Applied Cognitive Psychology*, *21*(5), 579-599.

Wallace, S. A., & Newell, K. M. (1983). Visual control of discrete aiming movements. The Quarterly Journal of Experimental Psychology, 35(2), 311-321.

Watson, C. S., & Kidd, G. R. (1994). Factors in the design of effective auditory displays. In Proceedings of the International Conference on Auditory Display (ICAD94) (pp. 293–303). Sante Fe, NM: ICAD.

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3(2), 159-177.

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.

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009, April). User-defined gestures for surface computing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1083-1092). ACM.

Woodworth, R. S. (1899). Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements*, *3*(3), i.

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. Pages 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.

Zaphiris, P. (2000). Depth vs Breadth in the Arrangement of Web Links. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 44(4), 453-456.

Zahariev, M. A., & MacKenzie, C. L. (2008). Auditory contact cues improve performance when grasping augmented and virtual objects with a tool.*Experimental brain research*, *186*(4), 619-627.

Zelaznik, H. N., Hawkins, B., & Kisselburgh, L. (1983). Rapid visual feedback processing in single-aiming movements. *Journal of motor behavior*,*15*(3), 217-236.

Zhao, S., Agrawala, M., & Hinckley, K. (2006, April). Zone and polygon menus: using relative position to increase the breadth of multi-stroke marking menus. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 1077-1086). ACM.

Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., Baudisch, P. (2007). Earpod: eyes-free menu selection using touch input and reactive audio feedback. CHI Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 1395-1404.

Appendix A

A.1 Experiment 4 t-value tables

A.1.1 Selection Time

Table A.1 t-values for paired comparisons for selection times

	AC	AD	Control	V	VAC
AD	5.398	--	--	--	--
Control	-12.131	-13.678	--	--	--
V	-11.746	-14.178	0.354	--	--
VAC	-15.360	-14.851	-1.927	-3.681	--
VAD	-14.985	-13.739	0.134	-0.271	3.309

A.1.2 Error

Table A.2 t-values for paired comparisons for adjusted error

	AC	AD	Control	V	VAC
AD	0.712	--	--	--	--
Control	9.225	9.177	--	--	--
V	-5.105	-6.030	-11.992	--	--
VAC	-5.526	-6.343	-12.139	-0.469	--
VAD	-4.070	-5.746	-11.585	1.238	1.987

A.1.3 Percent Accuracy

Table A.3 t-values for paired comparisons for percent accuracy

	AC	AD	Control	V	VAC
AD	-0.772	--	--	--	--
Control	-13.009	-14.160	--	--	--
V	3.181	4.558	18.104	--	--
VAC	3.638	4.587	17.577	-0.369	--
VAD	3.043	4.428	18.368	-0.771	-0.543

A.1.4 Undershoot and Overshoot

Table A.4 t-values for paired comparisons for difference in undershooting versus overshooting in endpoint positions.

	AC	AD	Control	V	VAC
AD	-1.952	--	--	--	--
Control	-6.277	-5.996	--	--	--
V	1.624	3.034	8.147	--	--
VAC	1.075	2.280	7.556	- 0.813	--
VAD	1.165	2.609	7.862	- 0.878	- 0.188

Table A.5 t-values for paired comparisons for time difference between entering target and making final selection.

	AC	AD	Control	V	VAC
AD	3.326	--	--	--	--
Control	-8.113	-8.765	--	--	--
V	-5.242	-6.670	0.522	--	--
VAC	-7.842	-7.832	-1.140	-2.401	--
VAD	-6.860	-6.703	0.662	0.097	3.285

A.1.5 Throughput

Table A.6 t-values for paired comparisons for throughput

	AC	AD	Control	V	VAC
AD	-6.083	--	--	--	--
Control	6.012	11.180	--	--	--
V	17.105	23.938	11.908	--	--
VAC	21.478	26.303	14.308	2.173	--
VAD	18.260	20.825	9.684	-0.966	-2.192