TOWARDS AN IN-VEHICLE SONICALLY-ENHANCED GESTURE CONTROL INTERFACE: A PILOT STUDY

Jason Sterkenburg, Steven Landry, Myounghoon Jeon, and Joshua Johnson

Mind Music Machine Lab

Michigan Technological University,

1400 Townsend Ave. Houghton, MI, USA

{jtsterke, sglandry, mjeon, jocjohns}@mtu.edu

ABSTRACT

A pilot study was conducted to explore the potential of sonically-enhanced gestures as controls for future in-vehicle information systems (IVIS). Four concept menu systems were developed using a LEAP Motion and Pure Data: (1) 2x2 with auditory feedback, (2) 2x2 without auditory feedback, (3) 4x4 with auditory feedback, and (4) 4x4 without auditory feedback. Seven participants drove in a simulator while completing simple target-acquisition tasks using each of the four prototype systems. Driving performance and eye glance behavior were collected as well as subjective ratings of workload and system preference. Results from driving performance and eye tracking measures strongly indicate that the 2x2 grids yield better driving safety outcomes than 4x4 grids. Subjective ratings show similar patterns for driver workload and preferences. Auditory feedback led to similar improvements in driving performance and eye glance behavior as well as subjective ratings of workload and preference, compared to visual-only.

1. INTRODUCTION

Touchscreens in vehicles have increased in popularity in recent years. Touchscreens provide many benefits over traditional analog controls like buttons and knobs. They also introduce new problems. Touchscreen use requires relatively high amounts of visual-attentional resources because they are visual displays. Driving is also a visually demanding task. Competition between driving and touchscreen use for visual-attentional resources has been shown to increase unsafe driving behaviors and crash risk [1]. Driving researchers have been calling for new infotainment system designs which reduce visual demands on drivers [2]. Recent technological advances have made it possible to develop in-air gesture controls. In-air gesture 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. Research has shown that accuracy of surface gesture movements can be increased with addition of auditory feedback [3]. However, there are many unanswered questions surrounding the development of an auditory supported in-air gesture-controlled infotainment system: What type of auditory feedback do users prefer? How can auditory feedback be displayed to limit cognitive load? What type of menu can offer an easily navigable interface for both beginners and experienced users? More importantly, do these displays reduce the eyes-off-road time and frequency of long off-road glances? Does the system improve driving safety overall when compared to touchscreens or analog interfaces? These are among the many questions that we attempt to address in this project, of which, this study is a first step. This study describes our efforts to develop an in-vehicle sonically-enhanced gesture control interface. The development of the prototypes draws from research in movement science, human-computer interaction (HCI), and auditory display research to develop prototype that improves on the safety of touchscreen interfaces.

2. DRIVING

2.1. Multi-tasking in Vehicles

In-vehicle information systems (IVIS), such as navigation devices, mobile phones, and radios often require manual input from drivers. If a driver wants to use an IVIS, he/she must balance the demands of the driving task with the demands of using the IVIS. Multiple Resource Theory [4] models how the demands of multi-tasking influence the performance on each of the tasks being completed. 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 similar resources, then performance on one, or both tasks will suffer. Both driving and IVIS use are primarily visual-manual tasks. Multiple Resource Theory predicts that driving performance may suffer as drivers attempt to use IVISs, as long as those IVISs require visual-manual resources to use. Auditory feedback has potential to facilitate IVIS use by providing driver with information without introducing competition for visual resources. Indeed, auditory feedback has been shown to improve menu navigation in IVISs [e.g., 5].

2.2. Eye Glances and Driving

Not all off-road glances are equal in their impact on driving performance. Compared to normal, baseline driving, short glances away from the road pose little or no risk to driving safety. Long glances away from the road – 2 seconds or more – increase near-crash/crash risk by at least two times normal driving [6]. The National Highway Traffic Safety Administration (NHTSA) has developed guidelines for IVIS design that suggest limits for permissible visual demands of IVIS use [7] which state that a driver should be able to complete tasks while driving with glances away from the road of 2 seconds or less. These guidelines and principles informed the design and analysis of the pilot study and will inform future iterations of the prototype design and future evaluations of the prototype effectiveness.

3. MOVEMENT SCIENCE

3.1. Fitts' Law

Paul Fitts' first quantified a movement task's difficulty, known as the index of difficulty (ID) [8, 9]. The original Fitts' Law equations describe movement along one dimension (1).

$$ID = log_2\left(\frac{2A}{W}\right), \tag{1}$$

Here, A is the amplitude, or distance, from the start of the movement to the target and W is the target width. The Shannon Formulation of Fitt's law (2) is generally preferred now because of its improved fit to observations while still adhering to Fitts' Law and because it ensures a positive value for ID.

$$ID = log_2\left(\frac{2A}{W} + 1\right), \tag{2}$$

This equation can help us predict the difficulty of completing movement tasks in our different systems. For example, when comparing movements toward similarly positioned targets in the two different grid sizes, such as target A in the 2x2 grid and target A in the 4x4 grid (Figure 1), if the amplitude is 50 cm for both grids (approximately true), and the target size in the 2x2 grid is 12.6 cm and 6.3 cm in the 4x4 grid, then the calculated ID for the 2x2 is 1.79 and the ID for 4x4 is 2.5. This predicts that selecting targets on the 4x4 will be more difficult. We do not suggest that Fitts' Law provides a complete description of the nature of these complex cognitive, visual-manual search tasks, but it does give us a foundation from which to make simple predictions about relative difficulty of using systems with different target sizes.

3.2. Auditory Feedback and Fitts' Law

Fitts' Law, and most of the related work done in the area of movement science have assumed that feedback about movement was obtained through the visual and proprioceptive modalities [10]. Research has shown that proprioceptive cues alone lead to reduced accuracy in movement tasks [11]. Since the in-vehicle gesture interface is intended to be used by drivers who are simultaneously driving a vehicle, visual resources may not be available. Proprioceptive cues alone may be insufficient to aid in movement toward targets. It is currently unclear how other feedback modalities, like auditory or haptic, can be best utilized to facilitate visually-unaided movement tasks while minimizing workload and unnecessary system noise.

4. PILOT STUDY

4.1. Objectives and Hypotheses

The purpose of this study was to evaluate the impact of two major design features on driving performance and driver glance behavior: the size and number of target boxes, and the presence of auditory feedback.

Hypothesis 1: We hypothesized that the larger target sizes would reduce the secondary task difficulty and result in better driving performance (lower lane deviations) and eye glance behavior (fewer glances, less eyes-off-road time, fewer long glances) compared to smaller target sizes.

Hypothesis 2: We also hypothesized that auditory feedback would decrease secondary task difficulty and result in better driving performance and eye glance behavior compared to conditions without auditory feedback.

4.2. Participants

A total of seven participants were recruited from Michigan Technological University undergraduate psychology student pool. Among the participants one was male and seven were female.

4.3. Equipment

4.3.1 In-vehicle Sonically-Enhanced Gesture Control Interface

The in-vehicle gesture interface is 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. Data from the LEAP Motion is sent to Pure Data, a free, open-source, real-time graphical programming environment for audio and visual processing. Within the customized Pure Data program there are audio and visual displays generated from the LEAP Motion data. The LEAP Motion tracks the center of a user's palm and counts the number of visible fingers and relays that information to Pure Data, which contains a visual grid display (Figure 1).
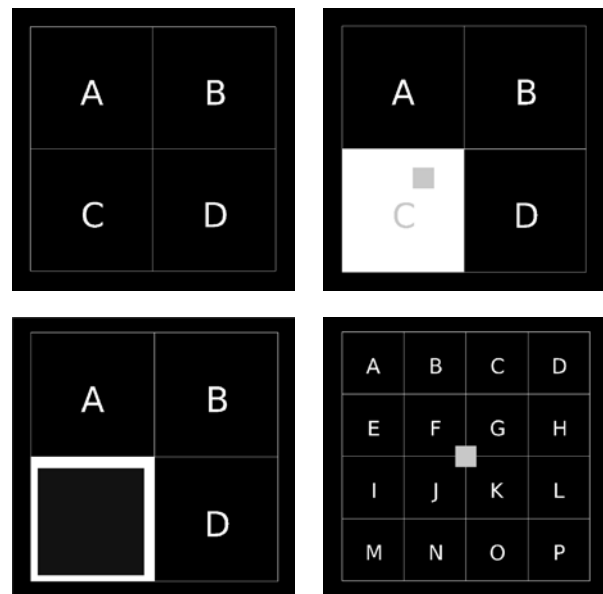

Figure 1: 2x2 grid (Top Left), 2x2 grid with visualization of hand position and highlighting box C (Top Right), 2x2 grid showing visualization of a selection (Bottom Left), and Graphical display of 4x4 grid with hand position (Bottom Right).

A graphic is displayed on a 1280x1024 monitor (Figures 1 & 2). The graphic shows a grid (2x2 or 4x4). Each box contains a letter. As the user holds his/her hand over the LEAP Motion, the visual display shows a box representing the position of his/her hand within the grid. If the center of the user's hand is within one of the boxes, that box is highlighted. For design concepts which have audio feedback, the same action will cue a text-to-speech .wav file for the letter in the box that is highlighted. Navigation through the system was

intended to be completed along a horizontal plane, with controls working analogously to a computer mouse. Target selection is dependent on the number of fingers visible to the LEAP Motion. If the system detects five fingers, then it will select the target which is highlighted at that moment. For the concept designs that have audio feedback, a selection action is followed by a confirmatory auditory icon 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.

4.3.2 Driving Simulator
A National Advanced Driving Simulator (NADS) MiniSim medium-fidelity driving simulator (Figure 2) was used for all driving scenarios. The driving scenario consisted of a single circuit through a residential area with many left and right curves. There were no other cars in the scenario. The simulator automatically records lane deviations and vehicle speed, along with many other variables.

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

4.3.3 Eye Tracking
Eye glance behaviors were recorded by a webcam placed on top of the visual display monitor. The eye glances were later coded by a researcher and placed into three categories based on the estimated length of the glance duration: short (<1 second), medium (1 second≤ t ≤2 seconds), and long (>2 seconds).

4.4. Experimental Design

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

- 2x2 grid with auditory feedback (2x2 VA)
- 2x2 without auditory feedback (2x2 V)
- 4x4 with auditory feedback (4x4 VA)
- 4x4 without auditory feedback (4x4 V)

4.5. Procedure

4.5.1 Training
Before driving in the simulator participants were introduced to the gesture prototype system. Initially, participants were shown the system and given no instruction in order to observe their first assumptions about how the system is used. A brief training period followed, in which participants were instructed to navigate with a closed fist and select by showing all five fingers. Practice trials were completed until the participant

was comfortable with the system. Next, participants were introduced to the driving simulator. Participants were told to drive in the right lane, and maintain a speed between 30-40 mph. The participants were given no instructions about how they should balance the demands of the two tasks.

4.5.2 Concept Systems
The order in which participants used the concept systems was randomized. A total of 32 selection tasks, evenly divided between target boxes, were completed for each concept 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 program.

4.5.3 Questionnaires
After completing all of the selection tasks, the participants were asked to stop the car and put it in park. During that time, the experimenter asked participants about his/her first impressions. Qualitative notes were taken regarding participants first impressions. Next, participants were asked several questions about their workload [12], including: mental demand, physical demand, performance, effort, and frustration using the electronic version of
NASA-TLX. This process was repeated for all four concept system designs.

4.5.4 Semi-structured Interview
Following completion of all concept system designs, a short interview was conducted to identify issues that participants noticed and to probe about experiences with various aspects of the system, including the target size and the presence of auditory feedback.

5.   RESULTS

5.1. Driving Performance

Speed data indicate that participants were generally capable of maintaining a speed between 30-40 mph, as instructed, while using each of the concept designs. Lane deviation data show a pattern indicating that participants' lane deviations were larger when using the systems with the smaller target sizes (4x4 grids) (Figure 3). Presence of auditory feedback appeared to have little or no effect on lane deviations.
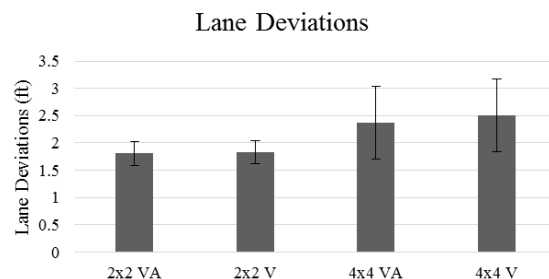

Figure 3: Mean lane deviations for each of the concept systems. Error bars denote 95% confidence intervals.

5.2. Eye Glance Behavior

Drivers made more frequent off-road glances for design concepts with smaller target sizes, and also for systems with no auditory feedback. This is true for all three glance durations (short, medium, long). The effect of both the target size and

the auditory feedback appears to be large. Target size and auditory feedback seem to act independently on glance durations, with no interaction occurring.
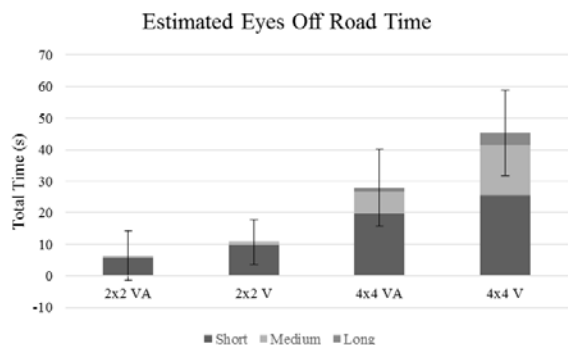

Figure 4: Cumulative eyes-off-road time for each of the concept systems. Error bars represent 95% confidence intervals.

5.3. Workload

NASA-TLX results show similar patterns for mental demand, effort, and frustration, each of which showed lowest scores for 2x2 VA, followed by 2x2 V, 4x4 VA, and 4x4 V. Perception of performance followed the reverse pattern, with the 2x2 VA grid resulting in highest perceptions of performance and the 4x4 V grid resulting in lowest perceptions of performance.

5.4. Semi-structured Interview

When participants were asked to rank-order their overall system preferences, they nearly unanimously favored systems in the following order: 2x2 VA, 2x2 V, 4x4 VA, 4x4 V. Two participants said that the auditory feedback was helpful for 2x2 grids but became more annoying than useful for 4x4 grids. Participants cited the ease of memorizing and acquiring the larger targets and the helpfulness of auditory cues (preview cues and confirmatory cues).

Researchers also observed that some participants initially attempted to control the device by moving vertically rather than horizontally. They stated that the vertical mapping was more intuitive to them. However, the current orientation mapping is used because movements tend to be faster along the x-plane than the y-plane [10]. Interestingly, participants would frequently move their hand down as they moved backwards, although no participants acknowledged conscious control over their downward movement.

6.   DISCUSSION AND FUTURE WORKS

The trends for all of the dependent measures indicate that larger target sizes, such as those in the 2x2 grids, lead to improved driving safety outcomes including lane deviations, eye glance frequency, eye glance duration as well as subjective measures of workload.

It is possible that the 2x2 grid is easier because the proprioceptive and/or peripheral visual information is sufficient to guide a person within the target range. Conversely, the smaller targets in the 4x4 grid may require additional visual information because the smaller targets cannot be acquired with proprioceptive information alone.

These results suggest that selection tasks with difficulty indices (ID) of 2.58 or higher should not be considered if the control space is located immediately in front of the in-vehicle center stack. Increasing target sizes and providing previewing and confirmatory auditory feedback can reduce secondary task difficulty and improve driving safety outcomes.

With the pilot study completed, we are developing custom software to allow us to test more refined designs. This software will come with configuration files allowing for a wider range studies. The new menu will be configurable to allow us to study the effects of variable menu layouts, different auditory displays for menu navigation (e.g., spoken titles, earcons, etc.), and record timing of participant actions. We will have predefined task sets defined within the software. Timestamps of each point of data from the start to the completion of the action will be recorded and will be later analyzed to better understand the relationship between a sonically-enhanced gesture controls and driving performance.

7.   REFERENCES

[1] W. Horrey, and C. Wickens, "In-vehicle glance duration: distributions, tails, and model of crash risk" Transportation Research Record: *Journal of the Transportation Research Board,* vol. 2018, pp. 22-28, 2007.

[2] P. Green, "Crashes induced by driver information systems and what can be done to reduce them," In *Sae Conf. Proc.* SAE; 1999, 2000.

[3] B. Hatfield, W. Wyatt, and J. Shea, "Effects of auditory feedback on movement time in a Fitts task," *Journal of Motor Behavior,* vol. 42, no. 5, pp. 289-293, 2010.

[4] C. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, pp. 159-177, 2002.

[5] M. Jeon, T. M. Gable, B. K. Davison, M. Nees, J. Wilson, and B. N. Walker, "Menu navigation with in-vehicle technologies: Auditory menu cues improve dual task performance, preference, and workload," *International Journal of Human-Computer Interaction*, vol. 31, no. 1, pp. 1-16, 2015.

[6] S. Klauer et al., "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data," (2006).

[7] P. Green, "Driver interface/HMI standards to minimize driver distraction/overload," (2008).

[8] P. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental Psychology,* vol. 47, no. 6, pp. 381, 1954.

[9] P. Fitts, and J. Peterson, "Information capacity of discrete motor responses," *Journal of Experimental Psychology,* vol. 67, no. 2, pp. 103, 1964.

[10] J. Medina, S. Jax, and H. Coslett, "Two-component models of reaching: Evidence from deafferentation in a Fitts' law task," *Neuroscience Letters,* vol. 451, no. 3, pp. 222-226, 2009

[11] S. Wallace, and K. Newell, "Visual control of discrete aiming movements," *The Quarterly Journal of Experimental Psychology,* vol. 35, no. 2, pp. 311-321, 1983.

[12] S. Hart, and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in Psychology,* vol. 52, pp.139-183, 1988.