Towards an In-vehicle Sonically Enhanced Gesture Control Interface

Jason Sterkenburg
jtsterke@mtu.edu

Michigan Technological University
CS5760
Topic Assignment #2

ABSTRACT

In-vehicle touchscreens provide many modern conveniences. However, touchscreen use also increases crash risk. A new type of in-vehicle gesture system is needed to improve driving safety while maintaining infotainment functionality drivers have come to expect. We propose an in-vehicle sonically enhanced gesture control interface. Relevant literature is introduced and results from a pilot study are described, followed by a discussion of potential future work.

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 unimodal 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].

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 I attempt to address in my dissertation, of which, this study is a first step. This study describes my efforts, in collaboration with undergraduate software engineers and other graduate student lab members to develop an in-vehicle sonically enhanced gesture control interface. The development the prototypes draws from research in movement science, human-computer interaction (HCI), and auditory research to develop prototype that improves on the safety of touchscreen interfaces.

The remainder of the introduction will be a review of relevant literature in the areas of driving, gestures, and movement science.

2. DRIVING

2.1. Multi-tasking in Vehicles

In-vehicle information systems (IVIS), like navigation devices, mobile phones, and radios require input from users in order to be used. 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 [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].

In order to improve driving safety, the National Highway Traffic Safety Administration (NHTSA) has developed guidelines for IVIS design that suggest limits for permissible visual demands of IVIS use [7]:

(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 [8]. 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 85th 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 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

Almost any potential IVIS design will require movement to use. As predicted by Multiple Resource Theory, movement task difficulty is among the factors that separate a safe IVIS from an unsafe one. Designing a system that requires easy movement tasks is especially important given that visual attention is primarily dedicated to the driving task. Fundamental laws of movement science help us to measure movement task difficulty and predict which tasks will be more difficult than others. Paul Fitts first quantified a movement task's difficulty, known as the index of difficulty (ID) (1) [9,10].

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

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 Fitts' 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.

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

The problem with those equations is that they account for movement along one dimension. Since this task requires movement in three dimensions, it requires a new formula, derived by Grossman and Balakrishnan [11].

$$ID_{WtEuc\theta} = log_2\left(\sqrt{\left(f_W\theta\left(\frac{A}{W}\right)^2 + \frac{1}{9.2}\left(\frac{A}{H}\right)^2 + f_D\theta\left(\frac{A}{D}\right)^2 + 1\right)}\right), \tag{3}$$

This equation accounts for three dimensions and also considers the movement angle – the angle at which the movement leaves the starting point. The terms A and W are the same as in the previous formulations of Fitts' Law and H is the height of the target and D is the depth of the target, and there are new constants: $fW(0°) = 0.211$, $fW(90°) = 0.717$, $fW(45°) = 0.242$, $fD(0°) = 0.194$, $fD(90°) = 0.312$, and $fD(45°) = 0.147$.

This equation should help us predict the difficulty of completing 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 50 cm for both grids (approximately true), and the target size in the 2x2 grid is 12.6 cm and 6.3 in the 4x4 grid then the calculated ID for the 2x2=1.79 and the ID for 4x4=2.5. This suggests that selecting targets on the 4x4 is more difficult.

Fitts' Law can also be used to predict the time it takes to perform movement tasks based on the ID of the movement task. In the context of complex navigation and search tasks, such as those performed by drivers using IVISs, Fitts' Law may be inappropriate for predicting movement times.

Movement science suggests that movement trajectories can be deconstructed into two phase: a ballistic phase and a corrective phase. The ballistic phase describes the initial, fast, smooth movement that has a rough trajectory towards the target. The ballistic phase is followed by the corrective phase which makes small corrections as discrepancies between current trajectory and target position are identified [15]. This phase is relatively slow are requires a lot of information, which is usually acquired through visual and kinesthetic channels.

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 [12]. Research has shown that proprioceptive cues alone lead to reduced accuracy in movement tasks [13]. Since the In-vehicle gesture interface is intended to be used by driver's who are simultaneously driving a vehicle, visual feedback 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 movement tasks while minimizing workload and unnecessary system noise.

4. MENU TYPES

The structure of a menu shapes the interaction method, and influences factors important in driving context, such as menu depth, time-to-complete tasks, and amount of visual or cognitive attention required to a complete task. There are two types of menu designs that are good candidates for in-vehicle use: point-and-click menus and marking menus. Point-and-click menus describe traditional desktop computer use. While mouse input is inadvisable, alternative sensors can provide positional inputs which can behave similarly to how a mouse works with a computer.

Marking menus may also be used to navigate IVIS menus. Marking menus use directional strokes as inputs. There are several subtypes of marking menus: polygonal, zonal, and radial. Polygonal marking menus are named because of the overall shape of target nodes and directional marks connecting them (Figure 1). Polygonal marking menus require extensive learning to master, which is prohibitive for in-vehicle use. Zonal marking menus map multiple controls onto unique stroke gestures by dividing the control space into multiple zones. When a similar gesture is performed in each of the zones, it controls a different function (Figure 2). Zonal marking menus are difficult to sonify because each zone contains two levels of information: a high-level description (e.g. “controls”) and a low-level description (e.g., “volume”). A radial marking menu describes a circlular graphic with menu options located along the outside edge of the circle. Strokes from the center of the circle toward a menu option will select that option (Figure 3). Radial menus are the most feasible option for in-vehicle implementation because they can be easily learned and sonified.

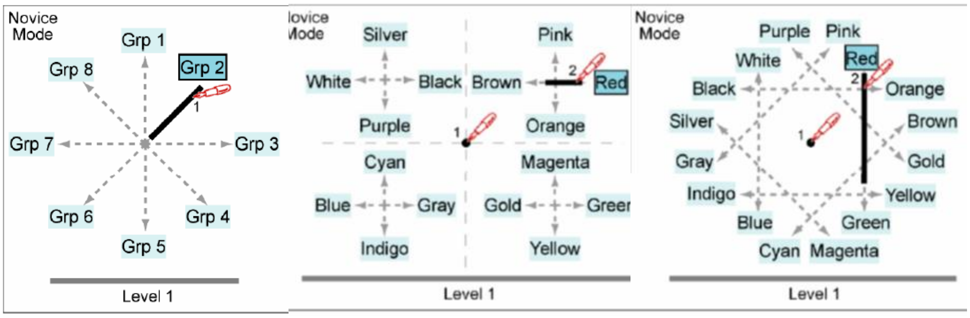

Figure 1: Adapted from [14]. (Left) Radial menu design. (Middle) Zonal menu. (Right) Polygonal menu.

Marking menus have specific advantages over point-and-click menus. The marking menu has no targets, it uses directional stroke gestures as input. Since there are no targets, the corrective phase of movement is totally eliminated. Ballistic movements are sufficient to select options in marking menus.

5. PILOT STUDY

5.1. Objectives

We conducted a pilot study to evaluate the impact of two major design features on driving performance and driver glance behavior: the size of target boxes, and the presence of auditory feedback. Only a point-and-click menu type was used for the first pilot.

Hypothesis 1: We hypothesized that the larger target sizes would reduce the secondary task difficulty and result in better driving performance 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.

5.2.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.

5.3. Equipment

3.1.1 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, detects the hand position of the driver. Data from the LEAP Motion is sent to Pure Data [16], 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 mapped onto a grid. 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 the grid display.

A visual graphic was displayed on a 1280x1024 monitor (Figures 2,3). 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 and 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. To reduce the number of false positive selections users were instructed to navigate to their target with a closed fist and select a target by opening their hand to show all five fingers. For the concept designs that have audio feedback, a selection action is followed by a confirmatory auditory earcon which contains two "raindrop" tones, the first low followed immediately by a second higher frequency note. This is intended to provide a positive indication of successful selection.

Auditory cues signal to participants which target to select. Each concept system has 32 target selection trials. Each target is sampled an equal number of times. The order of the auditory cues is randomly determined by the Pure Data program.

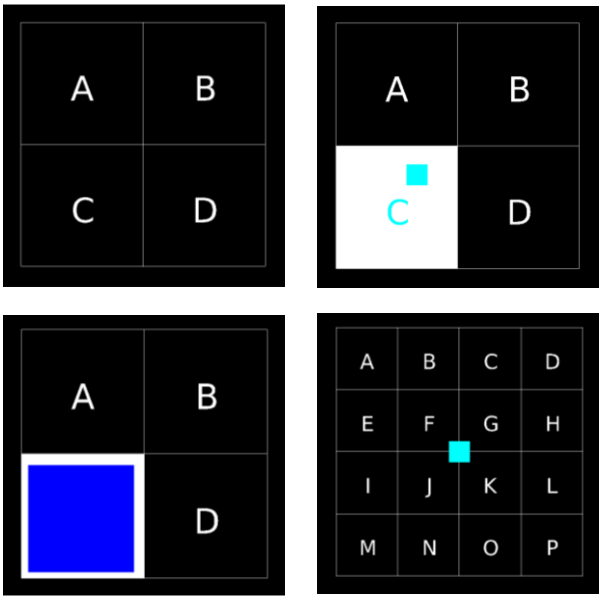

Figure 2: (Top Left) Graphical display of 2x2 grid, (Top Right) 2x2 grid with visualization of hand position and highlighting box C (Bottom Left) 2x2 grid showing visualization of a selection (Bottom Right) Graphical display of 4x4 grid with hand position.

3.1.2 Driving Simulator

A National Advanced Driving Simulator (NADS) Minisim medium fidelity driving simulator (Figure 3) 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 3: Driving simulator setup, visual display monitor with webcam, LEAP Motion.

3.1.3 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).

5.4. 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)

5.5. 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 options, 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 [17], including: mental demand, physical demand, performance, effort, and frustration using 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.

6. RESULTS

6.1. 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 4). Presence of auditory feedback appeared to have little or no effect on lane deviations.

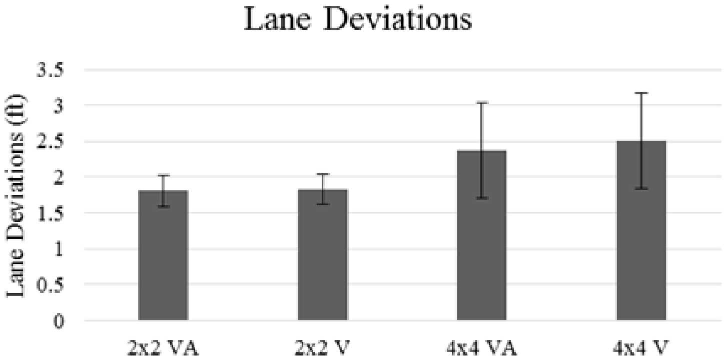

Figure 4: Mean lane deviations for each of the concept systems. Error bars denote 95% confidence intervals.

6.2. 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 5).

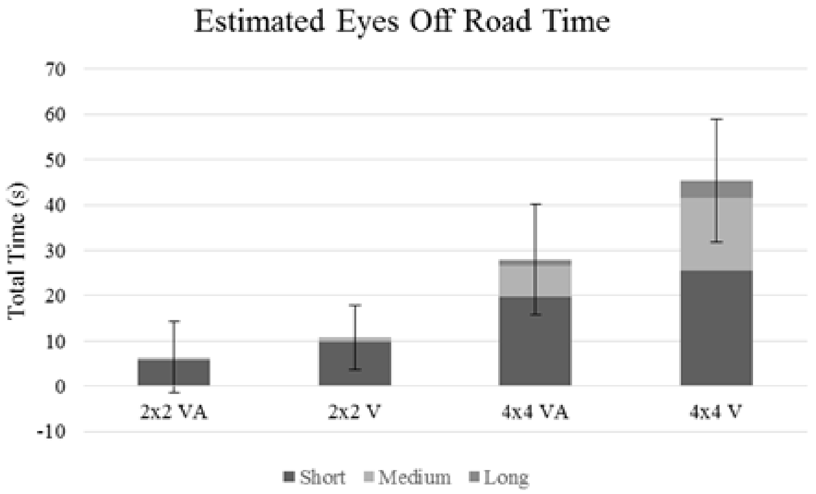

Figure 5: Cumulative eyes-off-road time for each of the concept systems. Error bars represent 95% confidence intervals.

6.3. 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.

6.4. 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 [11]. Interestingly, participants would frequently move their hand down as they moved backwards, although no participants acknowledged conscious control over their downward movement.

7. DISCUSSION AND FUTURE WORKS

Data from the pilot study suggest that target size has dramatic effects on ability to complete the tasks. Quantitative results did not highlight a similar effect from auditory feedback. However, researcher observations and questionnaire results suggest that auditory feedback reduced driver workload and frustration. The most important finding to emerge from the study was that movement tasks with an index of difficulty of 2.38 or greater should be eliminated from consideration.

Future studies will need to be conducted which compare a higher-fidelity prototype to a touchscreen while completing various prescribed tasks. Radial marking menus will need to be explored. They may provide an easier secondary task because they have a large, undefined target, unlike the point-and-click menu, such as the one described. It remains unknown how many menu options can exist in the menu while maintaining low error rates. This number is expected to be especially low considering that the selection gestures require only the ballistic phase of movement, which is faster but less accurate. It may be helpful to evaluate the effectiveness of an auditory-only display relative to visual-only and visual-auditory displays. Lastly, we need to define user-preferred gestures through a participatory gesture elicitation study. Gestures that are created by system designers are preferred less than user-generate gesture sets [18]. Defining user-preferred gestures is expected to improve the learnability of the system and the overall user experience of drivers.

Interestingly, movement tasks with indices of difficulty (ID) of 2.38 would ordinarily be performed with relative ease. Other variables may have influenced the movement task difficulty (separate from the visual search, and decision making subtasks). Two variables are considered. First, the selection gesture caused unintended variable movement of the cursor which caused errors. Second, visual tracking ability is diminished when driving. Participants were tracking their movements using interrupted focal glances and/or peripheral visual information. There may be a need to identify the practical range of IDs for visually-unaided movement tasks. Rather than defining a range of acceptably difficult movement tasks, another option for reducing secondary task difficulty would be to use a heads-up-display. A heads-up display may reduce secondary task difficulty by reducing the eccentricity of the visual targets.

Overall, the potential of an in-vehicle gesture control system remains largely undefined, but directions for future designs have emerged.

8. REFERENCES

[1] Horrey, William, and Christopher Wickens. "In-vehicle glance duration: distributions, tails, and model of crash risk." Transportation Research Record: Journal of the Transportation Research Board 2018 (2007): 22-28.

[2] Green, Paul. "Crashes induced by driver information systems and what can be done to reduce them." Sae Conference Proceedings p. SAE; 1999, 2000.

[3] Hatfield, Brent C., William R. Wyatt, and John B. Shea. "Effects of auditory feedback on movement time in a Fitts task." Journal of motor behavior 42.5 (2010): 289-293.

[4] Wickens, Christopher D. "Multiple resources and performance prediction." Theoretical issues in ergonomics science 3.2 (2002): 159-177.

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

[7] Boyle, L. N., et al. Text reading and text input assessment in support of the NHTSA visual-manual driver distraction guidelines. No. DOT HS 811 820. 2013.

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

[9] Fitts, Paul M. "The information capacity of the human motor system in controlling the amplitude of movement." Journal of experimental psychology 47.6 (1954): 381.

[10] Fitts, Paul M., and James R. Peterson. "Information capacity of discrete motor responses." Journal of experimental psychology 67.2 (1964): 103.

[11] Grossman, Tovi, and Ravin Balakrishnan. "Pointing at trivariate targets in 3D environments." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2004.

[12] Medina, Jared, Steven A. Jax, and H. Branch Coslett. "Two-component models of reaching: Evidence from deafferentation in a Fitts' law task." Neuroscience letters 451.3 (2009): 222-226.

[13] Wallace, Stephen A., and Karl M. Newell. "Visual control of discrete aiming movements." The Quarterly Journal of Experimental Psychology 35.2 (1983): 311-321.

[14] 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. Chicago

[15] Lank, Edward, Yi-Chun Nikko Cheng, and Jaime Ruiz. "Endpoint prediction using motion kinematics." Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, 2007.

[16] Puckette, M. S. (1996) Pure Data. Proceedings, International Computer Music Conference. San Francisco: International Computer Music Association, pp. 269–272.

[17] Hart, Sandra G., and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." Advances in psychology 52 (1988): 139-183.

[18] 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.