Information Architecture in Vehicle Infotainment Displays

by

Emily Gran

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2018 by the
Graduate Supervisory Committee:

Rob Gray, Chair
Russel Branaghan
Christina Carrasquilla

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

This study exmaines the effect of in-vehicle infotainment display depth on driving performance. More features are being built into infotainment displays, allowing drivers to complete a greater number of secondary tasks while driving. However, the complexity of completing these tasks can take attention away from the primary task of driving, which may present safety risks. Tasks become more time consuming as the items drivers wish to select are buried deeper in a menu’s structure. Therefore, this study aims to examine how deeper display structures impact driving performance compared to more shallow structures.

Procedure. Participants complete a lead car following task, where they follow a lead car and attempt to maintain a time headway (TH) of 2 seconds behind the lead car at all times, while avoiding any collisions. Participants experience five conditions where they are given tasks to complete with an in-vehicle infotainment system. There are five conditions, each involving one of five displays with different structures: one-layer vertical, one-layer horizontal, two-layer vertical, two-layer horizontal, and three-layer. Brake Reaction Time (BRT), Mean Time Headway (MTH), Time Headway Variability (THV), and Time to Task Completion (TTC) are measured for each of the five conditions.

Results. There is a significant difference in MTH, THV, and TTC for the three-layer condition. There is a significant difference in BRT for the two-layer horizontal condition. There is a significant difference between one- and two-layer displays for all variables, BRT, MTH, THV, and TTC. There is also a significant difference between one- and three-layer displays for TTC.

Conclusions. Deeper displays negatively impact driving performance and make tasks more time consuming to complete while driving. One-layer displays appear to be optimal, although they may not be practical for in-vehicle displays.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Introduction

As in-vehicle interfaces evolve to include a larger variety of tools and media, the ability for drivers to maintain safe driving practices becomes more and more at risk. Specifically, the use of such interfaces while driving can contribute to distracted driving (Sawyer et al., 2014). Therefore it is vital that the design of such interfaces facilitates the safest interaction and lowest levels of distraction for drivers possible. For that reason, it is important to examine the way information in infotainment centers is structured.

This study compares infotainment displays of with depths of either one, two, or three layers to determine how these differences in structure affect the driving performance of drivers using them. Shallow displays present multiple options per screen, with few levels for the user to advance through to find what they are looking for. In contrast, deep displays present less information on each screen and require users to advance through more levels to reach their goal. The depth or breadth of a display can have effects on working memory, which can contribute to distracted driving, and ultimately lead to safety risks (Burnett et al., 2013; Commarford et al., 2008; Jacko & Salvendy, 1996; Klauer et al., 2006; Young, Regan & Lee, 2009).

Distracted Driving and Safety

Distracted driving poses significant safety risks, not only to drivers operating vehicles, but also to all those who could be affected by their poor driving performance. Studies have examined the effect that driver distraction has on driving performance, and the results suggest that distraction while driving hinders safety. For example, the US Department of Traffic released a technical report, which concludes that distraction and lack of attention to the primary task of driving while operating a vehicle have been significant causes of accidents (Klauer et al., 2006).

Additionally, a meta-analysis conducted by Caird, Willness, Steel, and Scialfa, (2008), yielded results that fill in the gap between driving with distractions and causing car accidents. Caird et al. examined results from 33 studies with a total of about 2,000 participants, and concluded that using a cell phone while driving created decrements to driving performance, negatively impacting reaction time to critical events on the road. A mean increase of 0.25s in reaction time was found for all phone-related tasks that participants engaged in while driving. It was also mentioned that the actual decrement to driving performance would likely be greater when looking at the true behavior of participants when they drive and use their phones in their everyday life.

Furthermore, a second meta-analysis conducted by Caird, Johnston, Willness, Asbridge, and Steel, (2014), looking at texting and driving performance gathered data from 28 studies and 977 total participants. This meta-analysis concluded that reading texts while driving created a small decrement in driving performance, but reading and typing as well as typing alone had a substantial negative impact on eye movements, stimulus detection, reaction time, collisions, lane positioning, speed and headway. The task of typing or interacting with an interface impacts a driver's ability to pay adequate attention to the primary task of safe driving. This demonstration of how distracted driving and, more specifically, interacting with an interface while driving leads to degraded performance while operating a vehicle has important implications for drivers, but also for manufacturers of cars. The way cars are designed and interacted with can contribute to the level of distraction experienced while driving.

Working Memory Load and Distracted Driving

Uncovering the causes of driver distraction is therefore an important task when trying to ensure safe driving conditions. A study, (Ross et al., 2014), looking at the link between working memory load and driver distraction found that excessive demand on working memory is a significant contributor to driver distraction. The study examined working memory demand and its effects on performance of a lane change task. Participants were either exposed to working memory demands while driving or no working memory demands while driving and their driving performance was measured. The results concluded that higher demand on working memory led to poorer performance at the lane change task due to the distraction it caused.

Breadth vs. Depth and Working Memory

Other studies have confirmed the claim that deeper menu structures create more demand on working memory and, as a result, hinder performance. One experiment conducted by Commarford, Lewis, Smither, and Gentzler, (2008), required participants to take a pre-test assessing working memory capacity (WMC) and then asked them to complete a series of e-mail tasks using an interactive voice response system (IVR). The study consisted of two different groups, each group using the same IVR, but with different structures. In one condition, the IVR was designed with a broad, shallow structure, and in the other condition, the IVR was designed with a deep, narrow structure. The results concluded that there was a main effect of information structure for satisfaction and total time to complete a task, such that participants using the broader, shallower structured IVR completed tasks faster and experienced higher levels of satisfaction as a result of using the IVR than those in the deeper, narrower condition. There was also an interaction between total time to complete a task and WMC. Participants with lower WMC completed tasks with the broader menu structure in roughly the same amount of time as they did with the deeper menu structure. However, participants with higher WMC completed tasks much faster with the broader menu structure than the deeper one. There was also a main effect tasks completed, such that participants using the broader menu structure completed more tasks than those using the deeper menu. There was a significant main effect of WMC, such that those with higher WMC completed more tasks overall than those with lower WMC. However, WMC had no significant effect on tasks completed or satisfaction. These findings suggest that it takes longer to complete a task and fewer tasks can be completed when using deeper menu structures compared to broader ones. These findings also suggest that completing tasks with deeper menu structures takes more for people with lower WMC because the deeper displays put higher demand on working memory, which could lead to more distraction as a result.

Another study conducted by Burnett, Lawson, Donkor, & Kuriyagawa, (2013), examined the visual demand of in-vehicle interfaces. The study aimed to determine how the breadth or depth of a menu's structure could impact the visual demand of using the interface. Participants in this study were asked to use displays of varying breadths and depths (16x3; 8x4; 4x6; 2x12). The displays were also classified as structured or unstructured based on the way information was presented. Structured displays presented information options in alphabetical order, while unstructured displays presented information randomly, with no set guidelines. Visual demand was assessed by looking at glance frequency, glance duration, dynamic task time, and the amount of time that a participant's eyes were off the road. The results of this study showed that the lowest visual demand was found with broad, structured displays. There was no main effect of either organization or depth, but results did show an interaction between the two, such that participants favored breadth over depth for structured menus (lowest visual demand associated with 16x3 hierarchy). Conversely, for unstructured menus, participants favored compromise hierarchies (4x6; 8x4). Information in vehicles is often organized intentionally, grouping related items together and placing frequently used items at convenient locations in the display to the driver. Therefore, the results of this study support the idea that for structured menus, broad, shallow displays are advantageous compared to deep, narrow ones.

The perceived complexity of completing a task using an interface with either a broad or deep structure is another area that has been examined. Jacko and Salvendy, (1996), conducted a study in which participants used one of six different hierarchical menus varying in depth and breadth. The participants were asked to complete a series of tasks using these menus, and their performance was measured. Additionally, at the end of the experiment, they were asked to fill out a questionnaire to assess their perceptions of complexity pertaining to the tasks they completed. Results of this study showed that the deeper the structure of the menu was, the slower and less accurate participants were when completing tasks. Additionally, deeper displays were perceived to be more complex and difficult to use. This perception of complexity likely stems from the cognitive load of using deeper structured menus, because it is more difficult to remember the steps needed to reach the goal.

Optimal Menu Structure

These studies show that broader, shallower menu structures are advantageous compared to deeper, narrower ones. However, it is not always plausible for all information to be displayed on one screen at one time. For example, in user interfaces in vehicles, such as the infotainment system, there are pathways to get from broad options to more specific selections in order for the user to complete tasks. In these cases, it is important to determine how many levels are appropriate, as well as how many options should be included on each level.

In a study conducted by Miller, (1981), this issue is explored with regard to computer menus. In this study, participants used interactive computer menus with different depth and breadth to complete different tasks. They were then assessed on goal acquisition and accuracy. The results of this study showed that the best scores occurred for participants using menus with two levels and eight options per level. These criteria can serve as a baseline to measure against in the current experiment, with regard to in-vehicle interfaces and driving performance. The results of this study support the idea that broader, shallower structures are linked to better performance and are therefore more favorable than deep, narrow ones.

A study conducted by Kiger, (1984), further confirms these criteria. In this study, interfaces with several different structures were tested among participants. The structures varied in depth and breadth, as they did in the previous study. The researcher was examining how information architecture can affect information retrieval, and the results confirmed what Miller already discovered. In this study, it was found that two levels with eight or nine options per level yielded the best scores for information retrieval.

This baseline is not only similar to what Miller found, but is also supported by what scientists have already discovered about memory capacity. The results are consistent with how much information humans can store and retrieve from short term memory. Therefore, the structure of the interfaces in the current study will be influenced by these results. In the current study, interfaces will be designed with varying depth, but displays with more than one layer will include no more than eight or nine options per level. For displays with one layer, no memory effects should occur, because all options are displayed at all times.

Aims of the Current Study

The current study aims to uncover not only whether broad, shallow menu structure or deep, narrow menu structure is optimal for in-vehicle interfaces, but also to determine the how broad and shallow a menu can realisitcally be without negatively impacting performance by presenting too many items per screen and causing confusion. It has been concluded through previous research that broader menu structures are better overall, but this study tests the most shallow display structure possible and compares it to displays varying by only one or two layers. The depth of displays is limited to three layers deep at most, so all displays are fairly shallow in structure. Previous studies have looked at “shallow” or “deep” displays that differ by multiple layers, but this study will examine the impact of the shallowest possible display on driving performance, as well as increasing the depth of that display by only one or two layers to determine if even seemingly shallow display structures can be deep enough to negatively impact driving performance.

Display Structures. Based on the finding that two-layer structures are optimal, this study examines the driving performance of individuals interacting with menus whose depths range from one to three levels. One-level menus (Appendices B & C) contain 18 items displayed all at once in order to test the idea that a menu with more than nine items per level creates extraneous levels of distraction, negatively impacting performance. However, the 18 items are visually grouped into three categories of six items each in order to decrease cognitive load. There are two different one-layer menus; one displays items vertically and the other displays items horizontally. Figure 1 below illustrates a one-layer display with items organized vertically, as in condition 1. See Appendices B & C for full size and detailed images of the vertical and horizontal one-layer displays for conditions 1 and 2

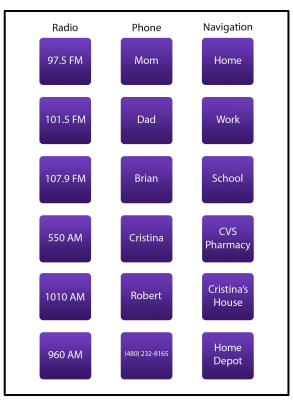


Fig.1 Vertical one-layer display with all items presented on one screen and organized into subcategories of "Radio," "Phone," and "Navigation."

There are also two two-level menus, displaying a maximum of six items at any given time. These menus are also organized either vertically or horizontally. Figure 1.1 below illustrates a two-layer display with items organized vertically, as in condition 3. Participants begin on the first screen and once a selection is made, they navigate to a second screen where they can make their final selection, as is shown below. See Appendices D & E for full size and more detailed images of the vertical and horizontal two-layer displays for conditions 3 and 4.

Fig 1.1 Vertical two-layer display with items broken into subcategories of “Radio,” “Phone,” and “Navigation.” Second screen allows for final item selection.

The three level menu (Appendix F) displays two or three items per screen and serves to test the effect of interacting with a deeper menu structure. The three level menu does not test the difference between vertical and horizontal display of items due to the fact that so few items are displayed at a time. Figure 1.2 below illustrates the three-layer menu used in condition 5. Participants begin on the first screen and make selections until they reach the third screen where they can make their final selection. See Appendix F for a full size and detailed image of the three-layer display for condition 5.

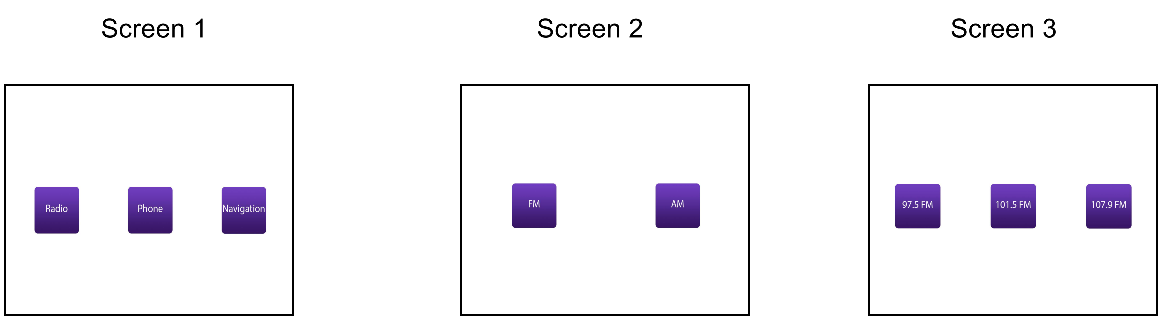


Fig 1.2 Three-layer display with items broken into categories of “Radio,” “Phone,” and “Navigation,” and then further broken down into subcategories on Screen 2. Screen 3 allows for final selection

Aside from the number of levels required to navigate these menus, the content displayed is identical for all menus. The purpose of this is to determine if driving performance is impacted more by the number of items on a screen at a time or the number of levels required to navigate through a menu.

The design and content of the displays is consistent in every way except for information architecture. Additionally, regardless of depth of menu structure, options that are displayed on each level are grouped into categories in order to avoid random presentation of items, as is recommended by Goubko and Danilenko, (2012). This consideration ensures that there is no confounding variable created by the confusion a user might experience if options were presented randomly. Random presentation of available options could cause the user to spend more time looking at the screen than it would typically take to complete a task in the real world.

Based on the findings of previous research in this area, the researchers predict the following: H_1: A broader menu structure will be associated with better driving and task performance than a deeper menu structure (Commarford et al., 2008; Burnett et al., 2013; Jacko & Salvendy, 1996; Miller, 1981). However, because this experiment limits the depth of all menu structures to three levels deep, it is predicted that, H_2: the two-level menu structure will be assocated with the best driving and task performance of all display depths. Additionally, it is predicted that, H_3: the menus with items organized vertically will be associated with better driving and task performance than those organized horizontally, because people read from left to right first, and then down. Therefore, it is predicted that it will be easiest for participants to first identify the category horizontally and then identify the item by reading down a column.

Driving performance measures (brake reaction time, mean time headway, and time headway variability) and task performance measures (time to task completion and success or failure of task) will be analyzed. Brake reaction time (BRT) is defined as the time to collision with the lead car when following less than three seconds behind and the driver steps on the brake. Mean time headway (MTH) is defined as the average distance in time between the participant's car and the lead vehicle. Participants are asked to maintain a two second time headway. If the time headway exceeds two seconds, it indicates that the participant is falling behind, and participants are told to speed up. Time headway variability (THV) is the variability in the previous variable, time headway. A higher value indicates that the participant is slower to react to changes in the speed of the lead vehicle. Time to task completion (TTC) is defined as the time it takes to complete a task from when the command is given to when the final selection is made. Failure of a task is defined as more than one step in the wrong direction to complete the task, or selection of the incorrect item. For example, if a participant tried to find an item that is listed under “Recent Calls” but mistakenly selects “Favorite Contacts,” this will not be defined as a failure. The participant could still go back and find the correct selection as long as he or she did not select the incorrect item as their final choice. However, if the participant selected the incorrect item or continues to search for the item in the wrong parts of the menu, this will be defined as a failure. No menu items are repeated, so selection of the incorrect item should not occur.

Methods

For this study, the independent variable is the structure of the display participants interact with. There are five experimental conditions, in which the same information is structured in varying levels of depth and breadth, as well as vertical and horizontal organization of menu items. There was also a baseline condition where participants drove without interacting with the infotainment display. This study is a repeated measures design so that all participants experience every level of the independent variable. Driver performance measures, including brake reaction time, mean time headway, and time headway variability, as well as task performance measures, including time to complete a task and success or failure of a task, are the dependent variables that were analyzed to determine the effect of interacting with the displays on the participant’s ability to focus on the primary task of driving.

The expectation of this study was that the one-level and three-level menu designs would have a more detrimental effect on driving performance than the two-level displays. This was predicted because previous research has shown that deeper displays create more working memory load because of the mental model of the information structure that needs to be conceptualized in order to remember how to complete certain tasks (Burnett et al., 2013; Commarford et al., 2008; Ross et al., 2014). Therefore, the three-level display should be more distracting than the two-level display. Additionally, it has been found that more than nine menu items presented at one time creates too much cognitive demand and leads to distraction (Kiger, 1984; Miller, 1981). Therefore, the one-level menus that present all 18 items at once should be more distracting than two-level menus that present six items at a time.

Participants

There were twenty participants recruited for this study, and data from nineteen participants was used. One participant was excluded due to an audio malfunction in the recording of the sessions for that participant. Each participant experienced five different conditions. The number of participants was determined using an a priori power analysis on previous research using similar measures of driver distraction (McNabb & Gray, 2016). Specific values used in the power analysis were: power=0.8, f=0.6. Participants for this study were recruited using the ASU Human Systems Engineering subject pool. All participants were 18 years or older and held a valid driver's license at the time of the study.

Materials

Driving simulator. A driving simulator was used for this experiment, and participants were asked to drive in it for the duration of their sessions. The driving simulator consists of the front half of a car mounted on a motion platform and wraparound screens that display the virtual environment which the participant is driving through. Please see Appendix A for a detailed description and photos of the simulator.

Displays. The conditions were designed so that there are five different display structures with varying breadths and depths. There were two displays with items presented in one layer, two displays with items presented in two layers, and one display with items displayed in three layers. For a detailed layout of each display condition, see Appendices B-F. The displays were created using Adobe Photoshop, as well as software called InVision and they were accessed via the InVision app on an iPad mini. The iPad mini was placed in the car where the center console would typically be, and participants were asked to interact with the iPad to complete tasks.

A video camera (Canon 70D) was used to record the driver's manual interactions with the different displays. There is a link to the videos of all participant driving sessions in Appendix J.

Procedure

Participants were given a brief overview of what the experiment entailed, and then were asked to fill out an informed consent. Participants were then given a demographic questionnaire including the number of years they have been driving and whether or not their current vehicle has a screen-based infotainment system in it. This questionnaire is included in Appendix G. Then, participants were directed to the simulator and asked to get inside the car and begin a practice drive. The practice drive was a five-minute drive in the simulator to become accustomed to it before beginning the experimental sessions. Participants completed five consecutive seven-minute driving sessions in the simulator, where they interacted with one of each of the five different displays per session.

In order to determine the order of the different levels of the independent variable for each participant, a partial Latin Squares counterbalancing scheme was used. Specifically, as shown in Appendix H, the orders were chosen to ensure that each display type occurred first and last an equal number of times.

Car following task. The car following task was identical to that used in several previous studies (Gray, 2011; Mohebbi, Gray, & Tan, 2009; Scott & Gray, 2008). Specifically, drivers followed a red lead car on a rural, two-lane road and were instructed to drive in their own lane and not pass the lead car. Drivers were instructed to maintain a 2.0 s time headway (TH) with the lead car. If the drivers followed too far behind the lead car, the words “Speed Up!” would appear in red text on the driver’s display. There was no analogous “Slow Down!” warning so that drivers were free to maintain any TH below 2.0 s. Drivers were given a five-minute practice drive (with no secondary task) to become familiar with the driving simulator and the car following task.

The lead car was programmed to unpredictably (to the driver) change speeds at variable intervals. The lead car traveled between 55 and 65 mph (with an average of 60 mph) with its speed determined by a sum of sinusoids. The lead car was programmed to make 8 unpredictable (to the driver) stops at a $-6 m/s^2$. The behavior of the lead car made it very difficult for the driver to predict when the lead car would speed up, slow down, or stop; creating multiple possible rear-end collision situations. Intermittent opposing roadway traffic was included to more closely simulate real-world rural driving conditions.

Secondary task: interacting with display. Participants began driving and were asked to complete nine tasks using the infotainment center throughout each seven-minute session. After each seven-minute session, participants were told to step out of the vehicle for a short break. During this time, the researcher switched the display to the next one in the sequence. Participants were also given a NASA TLX survey to complete in order to determine their perception of difficulty related to using each display. Then, the participant got back in the simulator and continued driving for another seven-minute session. This continued until all five levels of the independent variable had been experienced.

During each driving session, nine commands were given, telling participants to complete a task with the infotainment center. These tasks were: finding a certain radio station, selecting a certain destination, and making a phone call to a certain person. The order of tasks for each condition was randomly chosen, and all items were selected by participants at least twice throughout the entire experiment. The same tasks were given for each condition to every participant such that each participant interacting with Condition 1 completed the same tasks as every other participant that interacted with Condition 1. However, tasks and order of tasks were different for every condition. This was done so that participants' performance can be compared to other participants' performance directly for each condition, but the tasks were varied across conditions to avoid learning effects.

Each condition in this experiment was experienced only once in order to avoid the problem of learning effects. Additionally, the five consecutive driving sessions prevent regression effects that might occur if participants had to come back after a period of not using the simulator and get used to driving in it again. At the conclusion of all five sessions, participants were asked to fill out a final questionnaire to assess any preferences they may have had for any of the display types. This questionnaire is included in Appendix I.

Results

Separate one-way ANOVAs were conducted for each of the variables brake reaction time, mean time headway, time headway variability, and time to task completion. Paired samples t-tests were also conducted to examine the differences between one-, two-, and three-layer displays, regardless of item orientation (vertical vs. horizontal).

Brake Reaction Time (BRT)

There was a significant effect of brake reaction time. $F(1,18) = 19.44$, $p = 2.39e^{-05}$. As shown in Figure 2 below, the mean BRT for the 2-layer horizontal condition is significantly higher than any other condition.

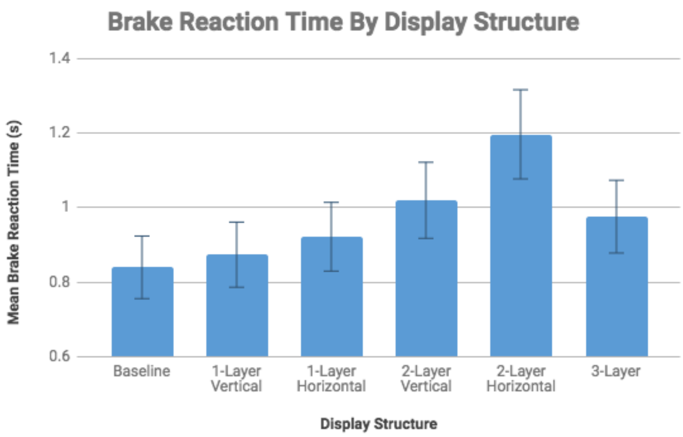


Fig. 2 Mean Brake Reaction Time across all display conditions and baseline condition, where participants did not interact with display while driving.

Mean Time Headway (MTH)

There was a significant effect of mean time headway. $F(1,18) = 9.42$, $p = 2.7e^{-03}$. As shown in Figure 2.1, the MTH for the 3-layer condition is significantly greater than any other condition.

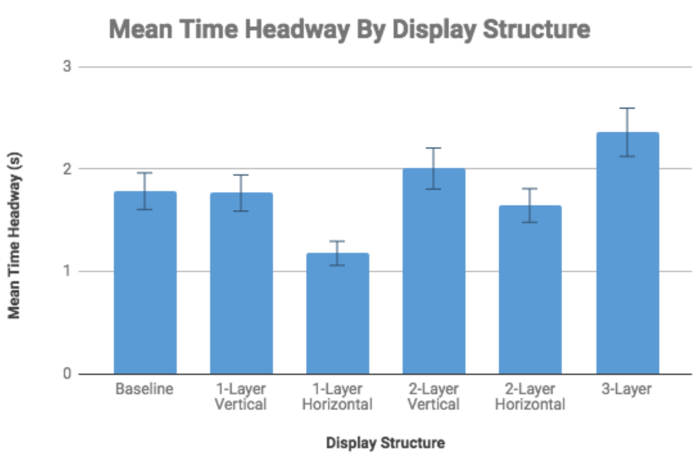


Fig. 2.1 Mean Time Headway across all display conditions and baseline condition.

Time Headway Variability (THV)

Although it was not significant, there was a marginally significant effect of time headway variability. $F(1,18) = 3.63$, $p = 5.92e^{-02}$. THV was higher in the 3-layer condition than any other condition, as shown in Figure 2.2.

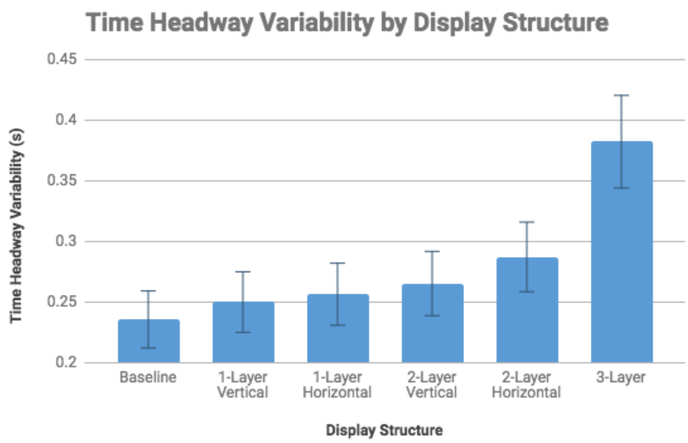


Fig. 2.2 Time Headway Variability across all display conditions and baseline condition.

Time to Task Completion (TTC)

There was a significant effect of time to task completion. $F(1,18)=25.35$, $p=2.33e^{-06}$. The TTC in the 3-layer condition was significantly higher than any other condition, as shown in Figure 2.3.

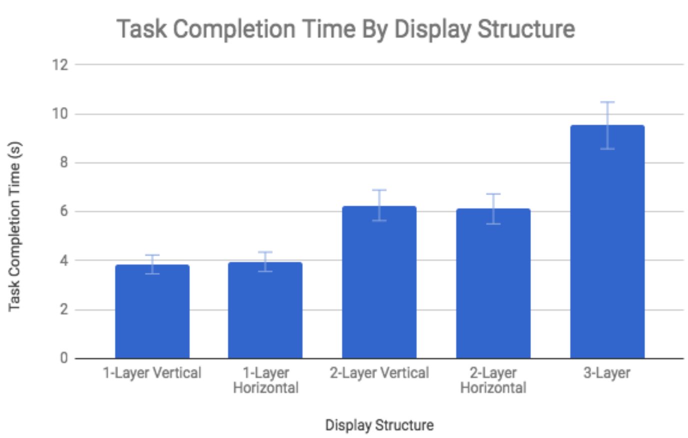


Fig. 2.3 Time to task completion across all display conditions.

Success or Failure of Task

Failure of a task was rare (only 5/855 tasks were failed). Therefore, failed tasks were omitted from the data analysis.

Comparing one-, two-, and three-layer displays

In order to further examine the effects of display structure on the variables of BRT, MTH, THV, and TTC, paired samples t-tests were conducted to compare one-, two-, and three-layer displays. Table 1 below shows the effects that were found.

There was a significant difference in BRT for one-layer (M=0.90, SD=0.14) and two-layer displays (M=1.11, SD=0.15); t(36)= -4.52, p=6.50e^{-05}. There was also a significant difference in MTH for one-layer (M=2.66, SD=0.13) and two-layer displays (M=2.81, SD=0.18); t(36)= -3.13, p=3.5e^{-03}. There was a significant difference in THV between one-layer (M=0.25, SD=0.03) and two-layer displays (M=0.28, SD=0.02); t(36)= -3.03, p=4.53e^{-03}. There was a significant difference in TTC between one-layer (M=3.92, SD=1.47) and two-layer displays (M=6.22, SD=2.76); t(36)= -3.21, p=2.79e^{-03}. There was also a significant difference in TTC between one-layer (M=3.92, SD=1.47) and three-layer displays (M=9.63, SD=6.86), t(36)= -3.55, p=1.11e^{-03}. There were no significant differences between one-layer and two-layer displays for any of the driving measures. There were also no significant differences for any of the variables between two-layer and three-layer displays. However, there was a marginally significant difference between two-layer (M=6.22, SD=2.76) and three-layer displays (M=9.63, SD=6.86); t(36)= -2.01, p=5.23e^{-02} for TTC.

Table 1 Significance of variables when comparing display depths

Layers	BRT	MTH	THV	TTC
1 vs. 2	**	*	*	*
1 vs. 3	NS	NS	NS	*
2 vs. 3	NS	NS	NS	MS*

** p<.0001
* p<.05
NS not significant
MS* marginally significant

DISCUSSION

The aim of this study was to examine the impact of in-vehicle infotainment display depth on driving performance. Specifically, the impact of completing tasks with one-, two-, and three-layer displays while driving was observed to determine at what depth a display becomes so distracting that it compromises driving performance. The results from the one-way ANOVAs for

MTH, THV, and TTC support H_1, which states that broader displays will be associated with better driving and task performance than deeper displays. The results from the one-way ANOVA for BRT also partially support H_1, because using a two-layer display hinders driving performance more than using a one-layer display. However, these findings do not fully support this hypothesis because the three-layer display was not associated with significantly higher BRT than the one-layer displays. Also, the results of the paired samples t-tests conclude that no significant differences were found between one- and three-layer or two-and three-layer displays for any of the driving performance measures. In fact, the finding that the two-layer horizontal display had the most negative impact on brake reaction time suggests that a two-layer display may have been more distracting than the three-layer display. This finding also directly contradicts H_2, which states that two-layer displays will be associated with the best driving and task performance.

This may be explained by the number of items displayed at one time combined with the requirement of navigating through a menu to make a final selection. Displaying all selectable items on one screen allows a user to scan the items and make a final selection without having to make any decisions aside from which item to select, which explains why the one-layer menu yielded the best performance. However, when items are broken into categories, decisions have to be made about which category one should choose in order to make a final selection. When this breakdown occurs in the two-layer menus, users are presented with six possible options to choose from to make their final selection, and this may be more confusing than the three items presented at the last screen of the three-layer menu. It could be easier for a user to further break down the information until there are fewer items that are closely related to one another. In this way, the user can treat each layer of the menu like a separate task that they can complete in smaller increments while driving. For example, if the command is to select 101.5 FM radio, users might find it easier to first make the selection “Radio,” return their eyes to the road, and then select “FM,” and return their eyes to the road once more before settling on 101.5 FM radio, rather than choosing “Radio” and being presented with six items that consist of both “FM” and “AM” stations. Seeing non-FM options on the screen may distract the user, causing confusion. This would also explain why the BRT for the three-layer condition was better than the two-layer horizontal condition. If the user has time to return his or her eyes to the road between quick selections, they might be more prepared for a sudden stop than if they spent more time on the second screen of the two-layer display trying to find the correct item among six options. It also explains the higher TTC for the three-layer condition compared to other conditions, because if participants treat each layer of the three-layer display like a separate task and continue driving in between, it would take longer to complete the entire task. Therefore, it would be optimal to display all items on one screen, but if this is not realistic, these findings suggest that it would be better to narrow down the items into smaller categories that are closely related in order for a user to easily navigate through the menu, rather than dividing items into broad categories.

The results of the analyses conducted show that deeper display structures negatively impact driving performance with regard to mean time headway and time headway variability. Drivers lagged further behind the lead car on average when interacting with the three-layer display than when interacting with any of the one- or two-layer displays. They also were less consistent about the distance they maintained behind the lead car when interacting with the three-layer display. This finding implies that interacting with deeper structured displays while driving may create more cognitive demand that detracts from the primary task of driving.

Additionally, it was found that drivers took the longest to complete tasks with the infotainment display when interacting with a three-layer display compared to a one- or two- layer display. This finding was expected, as drivers have to navigate further through the menu to make their final selection. The mean time to task completion for the five conditions were as follows: one-layer vertical mean= 3.85s; one-layer horizontal mean= 3.96s; two-layer vertical mean=6.26s; two-layer horizontal mean=6.12s; three-layer mean= 9.53s. A study conducted by Green (1999) concluded that secondary tasks completed while driving that require a driver to take his or her eyes off the road should take no more than fifteen seconds. All conditions in the current study allowed drivers to complete tasks in less than 15s on average. However, in the three-layer condition, there were 19 instances when participants took 15s or longer to complete a task. This number is much higher than all other conditions where five was the highest number of instances with a TTC of 15s or higher. While completing a task with an infotainment display in a vehicle, the driver is required to take his or her eyes off the road. Tasks that are more time consuming require the driver's attention to be diverted from the primary task of driving for longer than tasks that are not as time consuming. Therefore, deeper display structures lead to higher levels of distraction, because completing tasks with deeper displays takes more time than with shallower displays.

An interesting and unexpected finding is the effect the two-layer horizontal display had on BRT. Participants had significantly longer BRT when interacting with the two-layer horizontal display than any other display structure. It was expected that such an effect would be found in the three-layer condition, not a two-layer condition. This finding could be caused by the horizontal orientation of items, combined with multiple screens, which may have contributed to higher levels of distraction for the driver. This finding supports H_3, which states that vertically oriented menu structures will be associated with better driving performance than horizontally oriented menus. Due to the fact that people read from left to right and then top to bottom, the horizontally organized menu may be more confusing to use. The flow of the vertical menu displays broader category headings horizontally from left to right, and then each category opens up to reveal selectable items from top to bottom. This is consistent with the f-shaped visual search pattern that people tend to employ when interacting with interfaces (Nielsen, 2006). However, the horizontal menu first displays broader category headings from top to bottom, and then the selectable item fans out from left to right. This forces the user to utilize a visual search pattern opposite of the f-shape they may be used to. Because of this f-shaped visual search pattern, this finding may be related to how closely the menu structure resembles real-world applications. Menus used in web and mobile interfaces typically include drop-down menus where various related options appear beneath a broader category heading. The horizontal menu structure may not be as intuitive or easy to understand, which would explain why this condition is associated with the highest BRT.

Another possible explanation is that for the horizontal menu, the iPad mini that the display was presented on was turned horizontally, making the dimensions of the screen wider than that of the vertical display. When users interact with the horizontal display, they have to look further to the right to see the furthest-right menu options than they would with the vertical display.

This could draw their eyes further away from the road in front of them than the vertical display does.

Limitations

Technology. There were several limitations to this study that should be addressed when assessing the results. Driving and task performance were measured in this study, but it would have been valuable to assess driver distraction directly using technology like eye tracking. Eye tracking would have provided valuable insights about how much time participants spent looking at the road compared to how much time they spent looking at the display while completing tasks. A camera recording the participants' face while driving would have achieved similar results. It would be helpful to understand not only how long it takes to complete tasks, but also how much time was spent where participants' eyes were off the road while using displays of varying depths.

Another technological limitation of this study is the fact that the displays used were low-fidelity prototypes. For the purposes of designing and running this study in a timely manner, the displays have limited functionality. A display that could record and timestamp button presses would allow for more precise data analysis.

Ecological Validity. The low-fidelity nature of the displays used also presents limitations with regard to ecological validity. In-vehicle infotainment systems in cars, as well as mobile and web interfaces that participants are accustomed to interacting with are held to high standards of functionality and usability. The displays in this study are simple, and at times the lack of functionality may have been confusing for participants to use due to their expectations.

This can also be applied to the driving simulator itself. Driving in the simulator does not closely resemble real-world driving, and all participants have driving experience. Therefore, participants' behavior while driving in the simulator may not accurately represent their real-world performance. For example, although particpants are told to try not to hit the car in front of them, they may have been less motivated to avoid collisions in the simulator because they do not experience any direct repercussions like they would if they crashed a car in real life. It is also possible that participants felt they did not need to pay as much attention to the primary task of driving because they were not actually driving a car, and safety was not a major concern.

On the other hand, Hawthorne effects could have been at play as well. Participants were observed by an experimenter throughout each session, which may have impacted their performance. In a natural setting where there was no observer, participants may have been more likely to use their phones or engage in other distracting secondary tasks.

Participant Sample. Finally, the sample of participants in this study was limited to university students attending ASU who enrolled in this study for class credit. This sample was used for convenience, but a more representative sample may have impacted the results. It may have been useful to recruit participants with a specific amount of driving experience or of a specific demographic. For example, it may be useful to understand how infotainment display depth affects driving performance for individuals in the newest cohort becoming elligible to drive. It also might be interesting to examine the effect of infotainment display depth on older adults' driving performance.

Practical Implications and Future Research

The results of this study can inform in-vehicle infotainment display design. Although there were 18 menu items to choose from on one screen in the one-layer displays, which is more than twice the number recommended by Miller (1981), these displays yielded the best performance. Therefore, it may be useful for a display to present commonly-selected items all on one screen, rather than breaking items into categories and including multiple screens. More research should be conducted to determine the most common uses for infotainment system displays and how many unique tasks users perform on a regular basis.

Miller (1981) also recommended a two-layer display as the optimal menu structure for hierarchical computer menus, but the finding that the two-layer horizontal display in this study was associated with the highest average BRT suggests that a two-layer display may not be optimal for in-vehicle infotainment displays. More research should be done to compare two-layer menus with other depths, because the results of this study show that one-layer menus are associated with the better driving performance than two-layer menus. However, the results did not show a significant difference between two- and three-layer menus. There was also no significant difference between one- and three-layer menus, which suggests that two-layer menus, at least as they have been designed in this study, may not be as intuitive as the one- and three-layer menus used.

However, when comparing all conditions, the three-layer menu most negatively impacted MTH, THV, and TTC. A three-layer menu is still a fairly shallow menu, but this finding suggests that the depth of an infotainment display should be as shallow as possible. More research should be conducted to compare one-, two-, and three-layer menus to deeper-structured menus in order to determine at what point the average TTC exceeds 15 seconds, and is therefore dangerous to use while driving (Green, 1999).

It may also be useful to recreate this study using older adults as participants. A study conducted by Neena and Zimmer (2009) examined the impact of new technology in vehicles on older drivers and concluded that receptivity of an older adult to new technology relies largely on his or her concern for the problems that could be achieved by using the technology. Therefore, it may be less likely for an older adult to be interested in using an in-vehicle infotainment system unless there was a specific task they felt was necessary to complete while driving (i.e. making a phone call to a loved one). Therefore, research about what tasks older adults feel are important enough to complete while driving would give insight about important items to include on shallower layers of a display so they are more easily accessible to the driver.

Summary

This study examined the impact of in-vehicle infotainment display depth on driving performance and concluded that one-layer displays yielded the highest driving and task performance. Three-layer displays are associated with the lowest driving and task performance, with the exception of BRT, which was lowest in the two-layer horizontal condition. These findings suggest that shallower displays are more advantageous for in-vehicle infotainment displays, but it is unclear what the optimal structure should be. The results indicate that a one-layer display may be the least distracting for drivers, but there is no significant difference between the one-layer displays and the three-layer display. Therefore, more research should be conducted in order to determine whether three-layer menus are too deep to use safely while driving.

References

Burnett, G. E., Lawson, G., Donkor, R., & Kuriyagawa, Y. (2013). Menu hierarchies for in-vehicles user-interfaces: Modelling the depth vs. Breadth trade-off. *Displays, 34*(4), 241-249.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, *40*(4), 1282-1293.

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 and Prevention, 71*, 311-318.

Commarford, P. M., Lewis, J. R., Smither, J. A.-A., & Gentzler, M. D. (2008). A comparison of broad versus deep auditory menu structures*. Human Factors, 50*(1), 77-89.

Goubko, M. V., & Danilenko, A. I. (2012). Mathematical model of hierarchical menu structure optimization. *Automation and Remote Control, 73*(8), 1410-1423.

Gray, R. (2011). Looming auditory collision warnings for driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(1), 63-74.

Green, P. (1999). The 15-second rule for driver information systems. Intelligent Transportation Society of America Conference Proceedings (CD), Ann Arbor, Michigan, 1999. University of Michigan: Transportation Research Institute.

Jacko, J. A., & Salvendy, G. (1996). Hierarchal menu design: Breadth, depth, and task complexity. *Perceptual Motor Skills, 82*(3c), 1187-1201.

Kiger, J. (1984). The depth/breadth tradeoff in the design of menu-driven interfaces. *International Journal of Man-Machine Studies,* (20), 201-213.

D. Miller, The depth/breadth tradeoff in hierarchical computer menus. in: Proceedings of the Human Factors Society 25th Annual Meeting. Santa Monica, Human Factors and Ergonomics Society, CA, 1981.

Mohebbi, R., Gray, R., & Tan, H. Z. (2009). Driver reaction time to tactile and auditory rear-end collision warnings while talking on a cell phone. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(1), 102-110.

Neena L. Chappell & Zachary Zimmer, (2009) Receptivity to new technology among older adults, Disability and Rehabilitation, 21:5-6, 222-230.

Nielsen, J. (2006, April 17). F-shaped pattern for reading web content. Retrieved from https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/

S.G. Klauer, T.A. Dingus, V.L. Neale, J. Sudweeks, J. Ramsey, The impact of driver inattention on near-crash/crash risk: An analysis using the 100-Car naturalistic driving study data. US DOT technical report DOT HS 810 594, US Department of Transportation, Washington, DC, 2006.

Scott, J. J., & Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(2), 264-275.

Sawyer, B. D., Finomore, V. S., Calvo, A. A., and Hancock, P. A. (2014). Google Glass: A driver distraction cause or cure? *Human Factors, 56*(7), 1307-1321.

Ross, V., Jongen, E. M., Wang, W., Brijs, T., Brijs, K., Ruiter, R. A., & Wets, G. (2014). Investigating the influence of working memory capacity when driving behavior is combined with cognitive load: An LCT study of young novice drivers. *Accident Analysis & Prevention, 62*, 377-387.

APPENDIX A

DRIVING SIMULATOR

The DS-600c Advanced Research Simulator by DriveSafety™ will be used for this experiment. As shown in the photos below, the simulator displays the virtual surroundings via a wraparound screen. The simulator also includes a full-width Ford Focus vehicle cab and motion platform. Tactile feedback cues are provided using dynamic torque feedback from the steering wheel and vibration transducers under the driver's seat. The motion platform provides coordinated inertial cues during longitudinal acceleration and deceleration

APPENDIX B

ONE-LAYER VERTICAL CONDITION

In the one-layer vertical condition, all menu items are displayed in vertical columns, grouped by category. There are no hotspots, as all items are laid out on one screen. Once a participant makes a selection, there is no feedback or screen change.

APPENDIX C

ONE-LAYER HORIZONTAL CONDITION

In the one-layer horizontal condition, all items are displayed in horizontal rows, grouped by category. There are no hotspots, as all items are laid out on one screen. Once a participant makes a selection, there is no feedback or screen change. (*This photo was made smaller to fit on the page, but all buttons are the same size in the displays used for this experiment.)

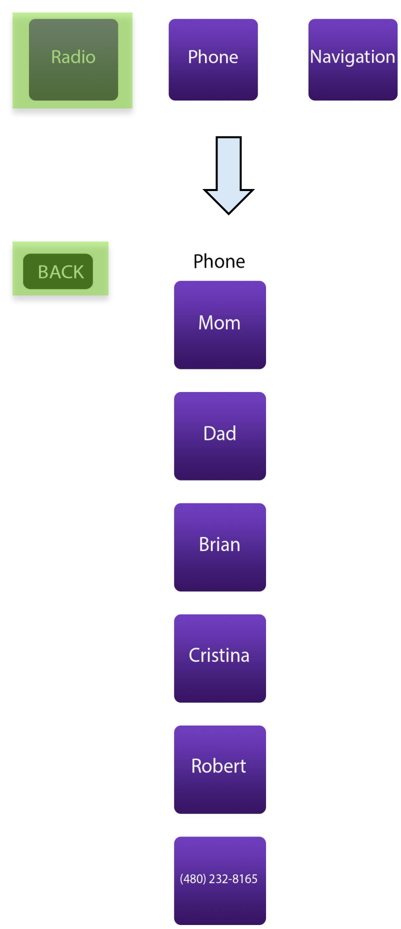
APPENDIX D

TWO-LAYER VERTICAL CONDITION

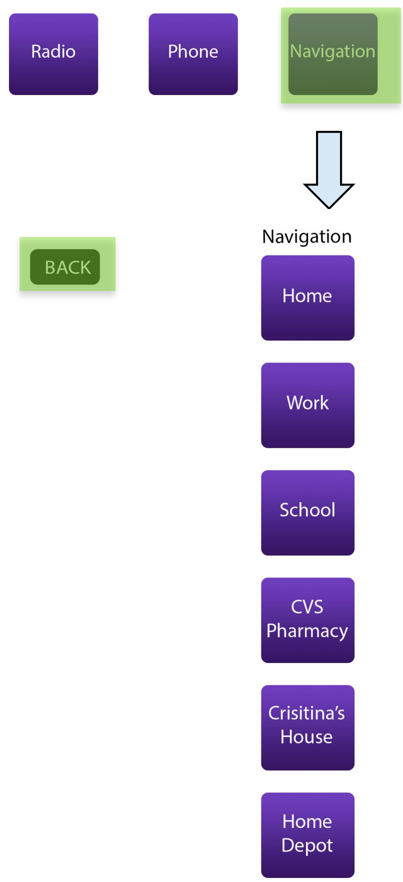
In the two-layer vertical condition, there are two levels to the menu. The main menu contains the category names, and then each category opens into the six menu items it contains. The menu items in this display are presented the same way that its one-level counterpart are displayed. This menu is the vertical organization, so if all three categories were opened at once, it would look like the one-level vertical menu. Hot spots that lead to subsequent pages are signified by a translucent green rectangle over an item that can be selected. (*Pictures not actual size.)

RADIO

PHONE



NAVIGATION

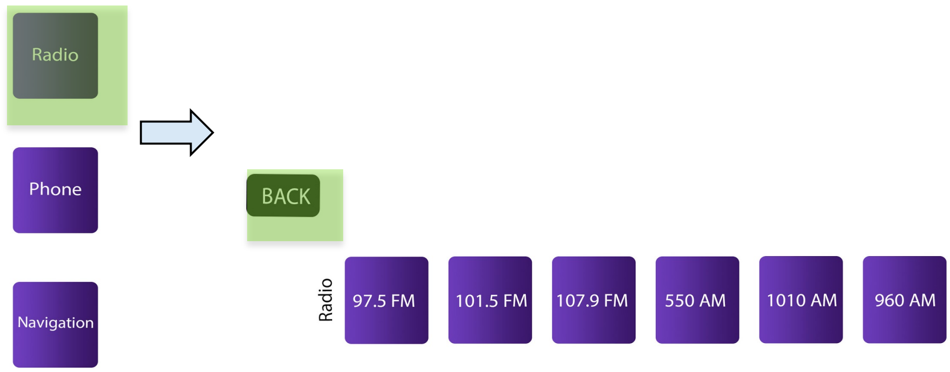

Radio
Phone
Navigation
BACK
Navigation
Home
Work
School
CVS Pharmacy
Crisitina's House
Home Depot

APPENDIX E

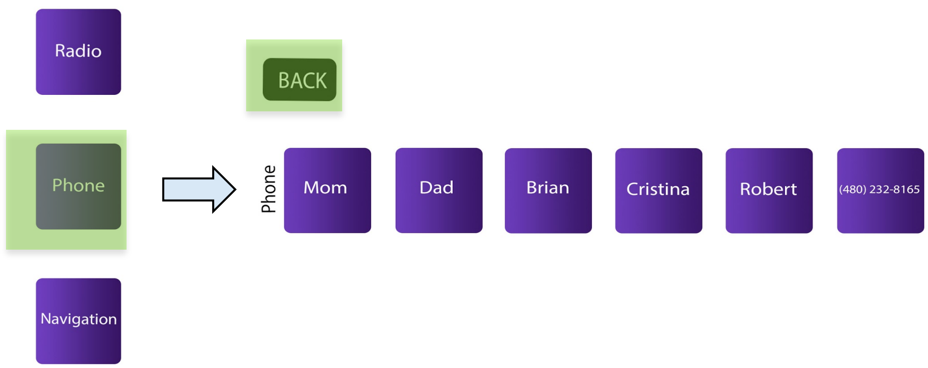
TWO-LAYER HORIZONTAL CONDITION

The two-layer horizontal menu is consistent with the design of the two-layer vertical menu, but main menu items are presented vertically and open horizontally to display all items. Hotspots are shown in green. Not actual size.

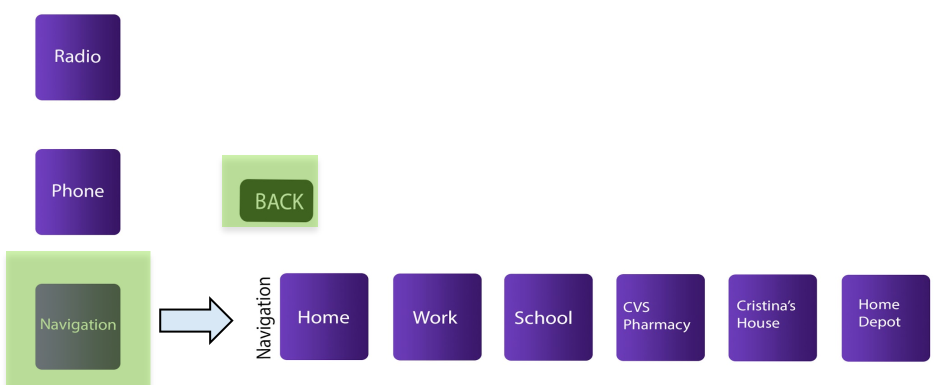
RADIO



PHONE



NAVIGATION



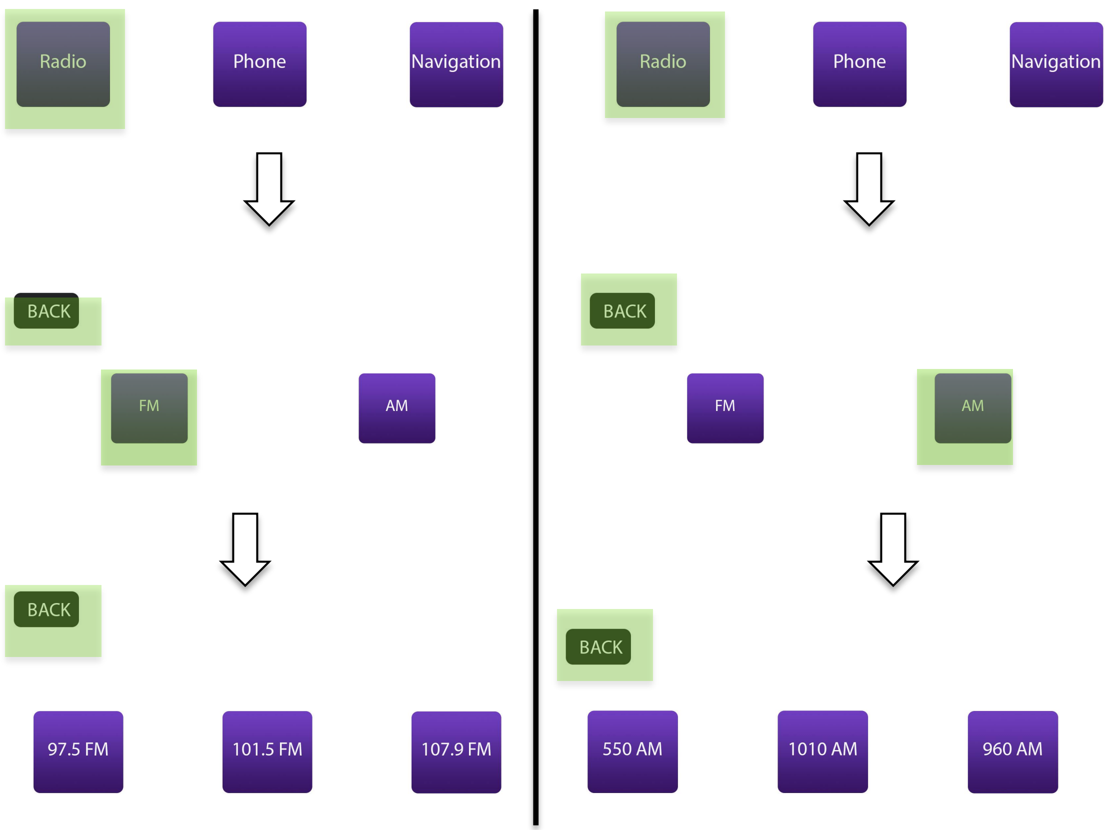
APPENDIX F

THREE-LAYER CONDITION

The three-layer condition is a three-layer menu in which all items are broken into three categories and then further divided into two more subcategories. There is no particular vertical or horizontal orientation to the display of items. Below, the menu structure flows for Radio, Phone, and Navigation are shown. Each category is broken into two subcategories, the flow of which are shown side-by-side. Hotspots are shown in green. (*Pictures not actual size.)

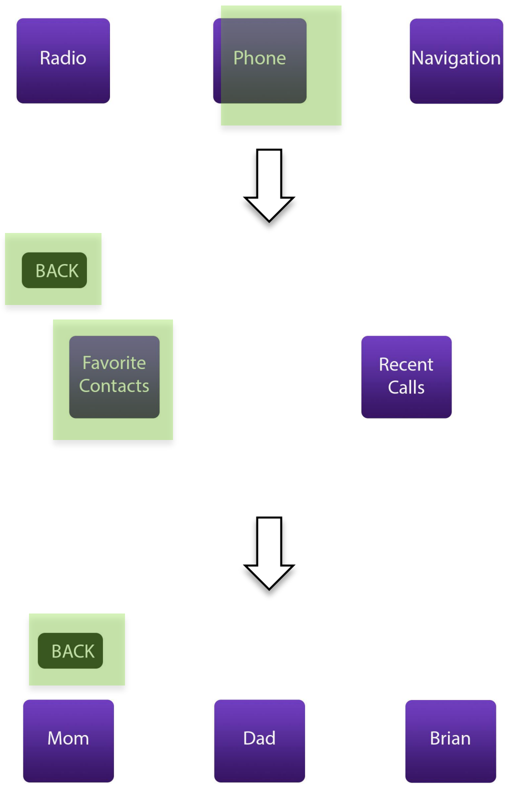
RADIO

<u>FM</u> <u>AM</u>

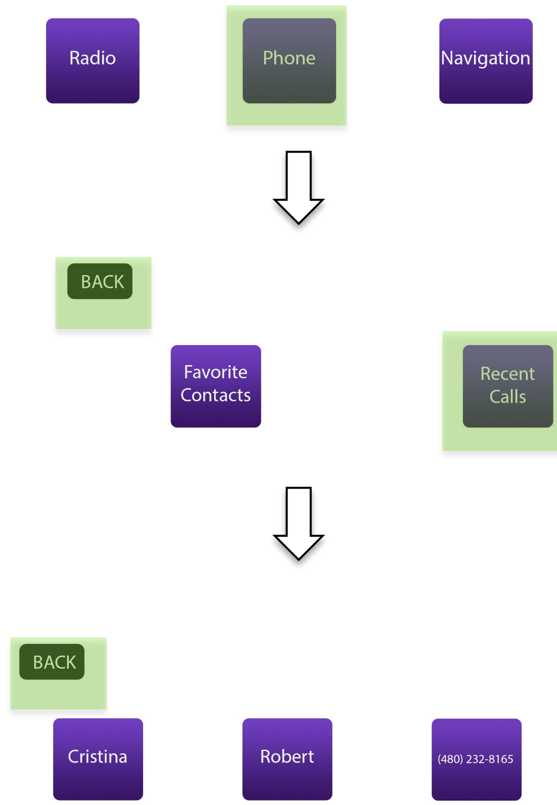


PHONE

Favorites

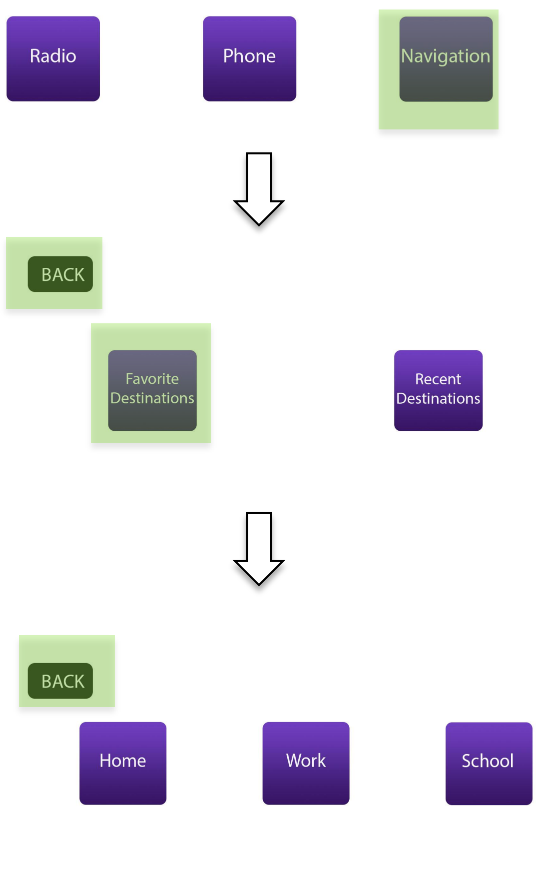


Recents



NAVIGATION

<u>Favorites</u> <u>Recents</u>



APPENDIX G

DEMOGRAPHIC QUESTIONNAIRE

1. What is your age? ____
2. What is your gender? (circle)
    a. Male
    b. Female
3. What is your current level of education?
    a. High school/GED
    b. Some college
    c. 2-year degree
    d. 4-year degree
    e. Master's
    f. Doctoral
4. How often do you use mobile devices?
    a. Multiple times per day
    b. At least once a day
    c. Every other day
    d. A few times per week
    e. Once a week or less
5. How long have you been driving a car?
    a. Less than a year
    b. 1-3 years
    c. 4-6 years
    d. 6-8 years
    e. 8-10 years
    f. Over 10 years
6. How often do you drive a car?

a. Every day
b. A few times per week
c. Once a week
d. Less than once a week

7. Do you currently drive a car?
    a. Yes
    b. No
8. If you answered “yes” to Question 7, does the car you drive have a screen-based infotainment system display in it?
    a. Yes
    b. No
9. If you answered “yes” to Question 8, how often do you use the infotainment system?
    a. Every time I drive
    b. Most of the time
    c. Occasionally
    d. Almost Never

APPENDIX H

PARTICIPANT TASK ORDER

1=vertical, one layer; 2=horizontal one layer; 3=vertical 2-layer; 4=horizontal 2-layer; 5=3-layer

Partial Counterbalancing:

1. 54213
2. 31245
3. 25134
4. 51324
5. 34152
6. 14253
7. 13542
8. 32514
9. 15324
10. 23451
11. 43152
12. 42513
13. 52431
14. 53142
15. 24513
16. 21345
17. 12435
18. 45321
19. 35421
20. 41235

APPENDIX I

PREFERENCE QUESTIONNAIRE

During this experiment, you completed tasks using various different displays. Please indicate below and preferences that you have for the different displays you used.

1. Some displays you interacted with displayed all the menu items on one screen, while others required you to click buttons to reveal menu items that were either two or three levels deep. Which type of display did you prefer?

a. One layer
b. Two layer
c. Three layer
d. No preference

2. If you chose A, B, or C for Question 1, please explain your preference below

____

____

3. Some displays you interacted with organized menu items of certain categories in vertical columns, while others organized menu items of certain categories in horizontal rows. What organization method did you prefer?

a. Columns
b. Rows
c. No preference

4. If you chose A or B for Question 3, please explain your preference below

____

____

5. Overall, was there a specific menu structure that you most preferred interacting with? If so, which one?

a. One layer, vertical
b. One layer, horizontal
c. Two layer, vertical
d. Two layer, horizontal
e. Three layer

APPENDIX J

PARTICIPANT DRIVING VIDEOS LINK

Below is the link to videos of all participant driving sessions on Google Drive.

https://drive.google.com/drive/folders/1w2LICUo2jmIgxsDkvzcEwnpY-k-fZ5sR?usp=sharing