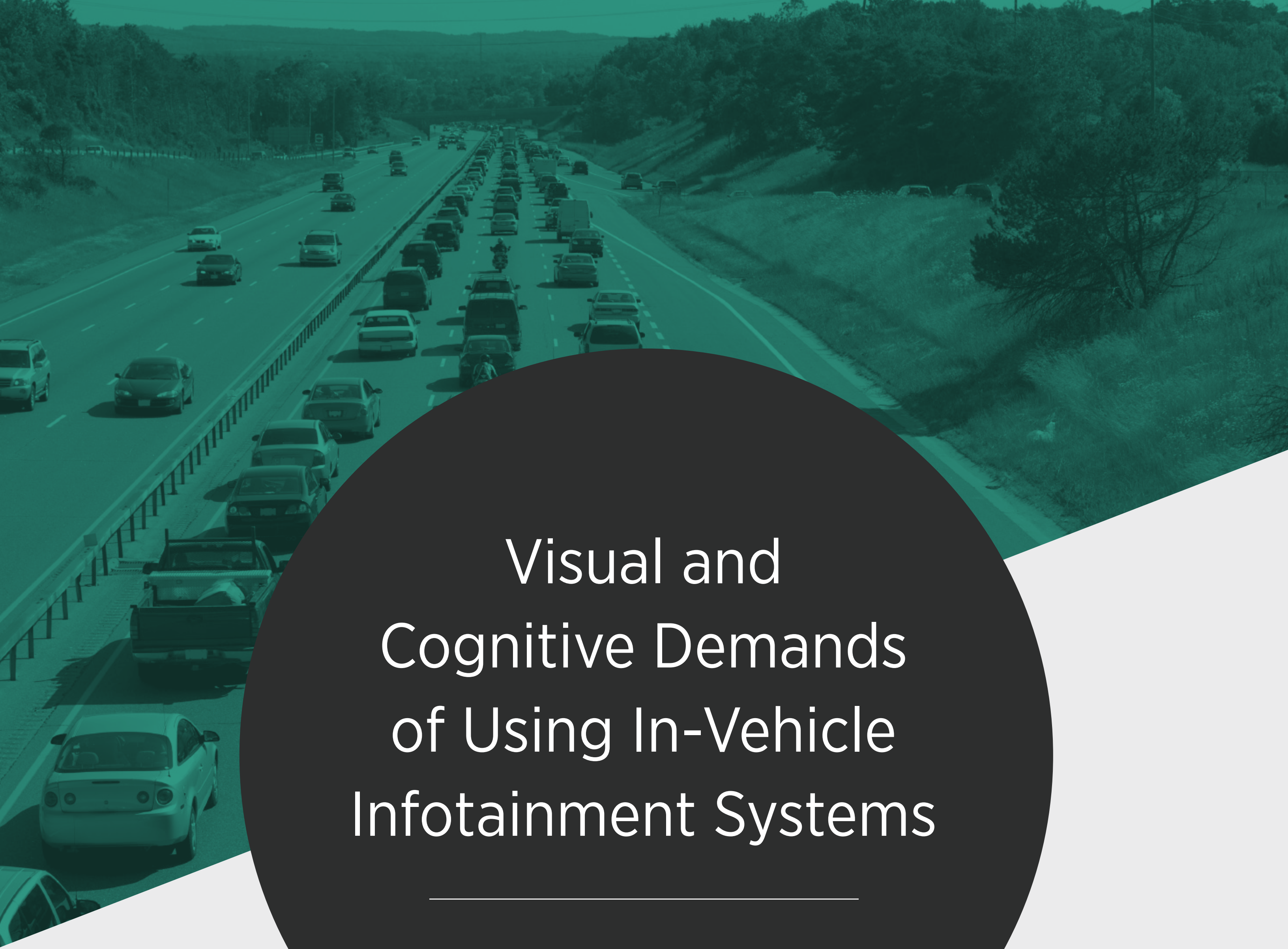

Visual and Cognitive Demands of Using In-Vehicle Infotainment Systems

September 2017

607 14th Street, NW, Suite 201 | Washington, DC 20005 | 202-638-5944

AAA Foundation for Traffic Safety

Title

Visual and Cognitive Demands of Using In-Vehicle Infotainment Systems
(*October 2017*)

Authors

David L. Strayer, Phd, Joel M. Cooper, Phd, Rachel M. Goethe, Madeleine M. McCarty, Douglas Getty, and Francesco Biondi, Phd
Department of Psychology
School of Social and Behavioral Science
University of Utah
Salt Lake City, UT

Foreword

The mission of the AAA Foundation for Traffic Safety is to save lives through research and education. One of major focus areas is understanding how emerging technologies can affect traffic safety. New infotainment and In-Vehicle Information Systems (IVIS) have greatly expanded in recent years, creating a wide array of tasks that motorists can perform while behind the wheel. Given the potential safety concerns, understanding how these new technologies impact drivers' workload and performance is paramount.

This report described the results of an on-road study looking at the visual and cognitive demand as well as the task completion time for a variety of infotainment tasks and interaction methods. Thirty 2017 model year vehicles, representing a wide range of manufacturers, were included in the study. This report and its outcomes should be a useful reference for automakers, developers of advanced IVIS, transportation agencies, public policy groups, researchers, as well as the general driving population.

C. Y. David Yang, Ph.D.

Executive Director
AAA Foundation for Traffic Safety

About the Sponsor

AAA Foundation for Traffic Safety
607 14th Street, NW, Suite 201
Washington, DC 20005
202-638-5944
www.aaafoundation.org

Founded in 1947, the AAA Foundation for Traffic Safety is a not-for-profit, publicly supported charitable research and education organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, individual members, AAA-affiliated insurance companies and other organizations or sources.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

2017 model-year automobiles provide a variety of features and functions that allow motorists to perform a plethora of secondary tasks unrelated to the primary task of driving. Many of these In-Vehicle Information Systems (IVIS) involve complex multimodal interactions to perform the secondary tasks. Surprisingly little is known about how these complex multimodal IVIS interactions impact a driver's workload. Given the ubiquity of these systems, the current research sought to address three interrelated questions concerning this knowledge gap. First, which task types are most distracting and what are the sources of distraction (e.g., visual/manual vs. cognitive)? Second, what is the workload associated with different modes of IVIS interaction (e.g., center stack, auditory vocal, center console)? Third, we directly compared the IVIS interactions supported by different OEMs to determine the bases for any differences in the workload associated with their use.

Depending on the availability of the IVIS features in each vehicle, our testing involved an assessment of up to four task types (audio entertainment, calling and dialing, text messaging, and navigation) and up to three modes of interaction (e.g., center stack, auditory vocal, and the center console). Three additional tasks were used to facilitate the assessment of the visual/manual and cognitive demands of the IVIS interactions. The first was the single-task baseline where participants drove the vehicle without any secondary-task interaction. The second, a cognitive referent task, was a concurrent auditory/vocal N-back secondary task that placed a high level of cognitive demand on the driver without imposing any visual demands (Mehler, Reimer, & Dusek, 2011). The third, a visual referent task, was a concurrent Surrogate Reference Task (SuRT, ISO TS 14198), which placed a high level of visual/manual demand on the driver. The N-back and SuRT tasks were adjusted so that they were equivalent in difficulty when compared with the single-task baseline (i.e., Cohen's d was 1.423 and 1.519, respectively).

For each of the 30 vehicles tested in the current research, 24 participants were evaluated as they drove on a residential road with a posted speed limit of 25 mph. After familiarization with the road, the vehicle, the IVIS tasks, and the modes of interaction, testing commenced, with the order of the testing conditions counterbalanced across participants. A number of performance measures were obtained while participants performed the tasks including primary-task measures, secondary-task measures, and subjective measures. Primary-task measures included GPS data and video recording of the driver and the driving environment. Secondary-task measures were obtained using two variants of the Detection Response Task (DRT, International Organization for Standardization #17488). Subjective measures were obtained at the end of each condition using the NASA Task Load Index.

The data collected from each participant provided a measure of cognitive demand, a measure of visual/manual demand, a subjective workload measure, and a measure of the time it took to complete the different tasks. These metrics were evaluated separately and also combined to provide an overall demand score for the different tasks, modes of interaction, and vehicles. These metrics were standardized relative to the high demand cognitive referent (i.e., the N-back task had a rating of 1.0) and the high demand visual referent (i.e., the SuRT task had a rating of 1.0). Using this integrated metric, task types, modes of interaction, and vehicles that had a rating between 0.0 (the demand associated with the single-task baseline) and 1.0 were easier than the high-demand referent and those with ratings greater than 1.0 were harder than the high-demand referent. This procedure also provided a metric for directly comparing different tasks, different modes of interaction, and different vehicles.

Our analysis found that the IVIS *task types* differed in terms of visual and cognitive demand, with the audio entertainment task type being equivalent to the calling and dialing task type (the two most universal of IVIS tasks available in all 2017 model-year automobiles we tested). Text messaging, an IVIS feature found in 22 out of 30 vehicles we tested, was associated with a significantly higher level of demand than the former task types. Most demanding of all was destination entry for navigation, an IVIS feature that was available in 12 out of 30 of the vehicles we evaluated. The navigation task type had an overall demand that was more than two times that of the high demand referent.

Second, we found that the overall workload associated with each *mode of IVIS interaction* was greater than the high workload referent. Interactions using the center stack were significantly less demanding than auditory vocal interactions, which were less demanding than center console interactions. Interestingly, using voice-based commands to control IVIS functions resulted in lower levels of visual demand than the SuRT task. However, the benefits of reduced visual demand were offset by longer interaction times. Auditory vocal interactions took significantly longer than any other IVIS interaction (an average of 30 seconds in our testing).

Finally, our analysis found surprisingly large differences between *vehicles* in the overall demand of IVIS interactions. Seven of the 30 vehicles received an overall rating significantly below 1.0 (i.e., a moderate level of overall demand). Eleven of the 30 vehicles received a score that did not differ from the high demand referent (i.e., a high overall demand score). Twelve of the 30 vehicles scored significantly above the high demand referent (i.e., a very high overall demand score). On the whole, vehicles in the latter category tended to have higher levels of demand on cognitive, visual, and subjective measures as well as longer interaction times.

The vast majority of the IVIS features and functions in the vehicles we evaluated were unrelated to the task of driving (or, in the case of destination entry to support navigation, could have been performed *before* the vehicle was in motion). Many had cumbersome human-machine interfaces with design inconsistencies that lead to high levels of workload. In fact, many IVIS interactions were associated with high levels of cognitive and visual demand with long interaction times. For example, 83% of the vehicles with a very high overall demand offered destination entry for navigation while the vehicle was in motion, an IVIS task we found to produce high levels of workload.

Our research provides empirical evidence that the workload experienced by drivers systematically varied as a function of the different tasks, modes of interaction and vehicles that we evaluated. Our objective assessment suggests that many of these IVIS features are too distracting to be enabled while the vehicle is in motion. This is troublesome because motorists may assume that features that are enabled when they are driving are safe and easy to use. Greater consideration should be given to what interactions *should* be available to the driver when the vehicle is in motion rather than to what IVIS features and functions *could* be available to motorists.

TERMS AND DEFINITIONS[1]

Center console – The center console is located between the driver and passenger front seats. Examples of center console interactions include a rotary dial that allows drivers to scroll through menu items and a writing pad where drivers use their finger to write out commands.

Center stack – The center stack is located in the center of the dash to the right of the driver. An LCD display is used to present textual and/or graphical information. Center stack systems often include a touch-screen interface to support visual/manual interactions so that drivers can select an option and navigate menus by touch and/or use slider bars to scroll through options displayed on the screen. With some vehicles, the selection of options may be made with manual buttons surrounding the touch screen.

Cohen's d – An effect size estimate derived by a standardized difference between means. A Cohen's d value of 0.2 reflects a small effect size, a value of 0.4 reflects a medium effect size, and a value of 0.8 reflects a large effect size.

Cognitive demand – The cognitive workload associated with the performance of a task. This would include perception, attention, memory, and decision-making processes. In this report, we refer to the cognitive demand associated with performing IVIS task types with different modes of interaction when the vehicle is in motion.

Cognitive referent task – The N-back task (see below) served as the cognitive referent task in the current research.

Overall demand – Total visual, auditory, cognitive, or physical resources required of the driver to accomplish the primary driving task and interact with an in-vehicle infotainment system in a dual-task setting.

Distraction potential – The potential distraction associated with secondary-task engagement. This potential may not be realized if drivers regulate their secondary-task interactions to periods when the vehicle is not in motion.

Driver distraction – The diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving.

DRT – The Detection Response Task (DRT) is an International Standards Organization protocol (ISO 17488, 2015) for measuring attentional effects of cognitive load in driving. In this research, a vibrotactile device emitted a small vibration stimulus, similar to a vibrating cell phone or an LED light stimulus changing color from orange to red. These changes cued the participant to respond as quickly as possible by pressing the microswitch attached to a finger against the steering wheel. DRT reaction time increases and hit rate decreases as the workload of the driver increases.

Dual task – Two tasks performed concurrently, typically the primary driving task plus a secondary task.

Evaluation – A procedure for assessing the effects of an interaction.

Impairment – The degraded driving performance associated with secondary-task interactions. This includes compromised hazard perception, slower brake reaction time, degraded lane keeping, etc.

[1] Some of these terms, definitions, and abbreviations were taken directly from ISO/TS 14198, Regan, Hallett and Gordon (2011), and NHTSA (2013).

Impairments can be assessed using measures of driving performance or through changes in secondary tasks, such as the DRT.

In-vehicle information system (IVIS) – The collection of features and functions in vehicles that allow motorists to complete tasks unrelated to driving while operating the vehicle. The IVIS features we tested involved up to four task types (audio entertainment, calling and dialing, text messaging, and navigation) and up to three modes of interaction (e.g., center stack, auditory vocal, and the center console).

Method – High-level approach to an assessment, based on theory and principles, which implies an underlying rationale in the choice of assessment techniques.

Metric – Quantitative measure of driver behavior independent of the tool used to measure it.

Linear mixed effects model – We compared the likelihood ratio of the full linear mixed effects model to a partial linear mixed effects model without the effect (e.g., Task, Mode, Task by Mode, Vehicle) to determine if the effect in question accounted for a significant proportion of variance.

NASA TLX – A questionnaire-based metric assessing the subjective workload of the driver. The TLX assesses mental demand, physical demand, temporal demand, performance, effort, and frustration.

N-back task – The N-back task presented a prerecorded series of numbers ranging from 0 to 9 at a rate of one digit every 2.5 seconds. Participants were instructed to say out loud the number that was presented two trials earlier in the sequence. The N-back task places a high level of cognitive demand on the driver without imposing any visual/manual demands.

Performance – The behavior demonstrated by a driver performing the driving task or a related task.

Primary driving task – Activities that the driver must undertake while driving including navigating, path following, maneuvering, and avoiding obstacles.

Reference task – Type of task used for the purpose of comparing different tests or test results across vehicles or systems.

Single task baseline – When the driver is performing the primary driving task (i.e., driving) without the addition of workload imposed by IVIS interactions.

Secondary task – A non-driving related additional task.

Secondary task demand – The aggregate of cognitive, visual, and manual demands required by a non-driving task.

SuRT task – The variant of the Surrogate Reference Task (SuRT, ISO TS 14198) used in this report required participants to use their finger to touch the location of target items (larger circles) presented in a field of distractors (smaller circles) on an iPad Mini tablet computer that was mounted in a similar position in all the vehicles. The SuRT task places a high level of visual/manual demand on the drivers because they must look at and touch the display to perform the task. The SuRT task served as a referent for the visual/manual demands associated with performing IVIS interactions.

Task – The process of achieving a specific and measurable goal using a prescribed method.

Task interaction time – The time to complete a task. Task interaction time was defined as the time from the moment participants first initiated an action to the time when that action had terminated and the participant said, “done.”

Task types – Tasks were categorized into one of four task types: Audio entertainment, calling and dialing, text messaging, and navigation, depending on vehicle capabilities. These task types were completed via different modalities equipped in each vehicle (i.e., touch screen, voice recognition, rotary wheel, draw pad, etc.) for each interaction.

Total task time (TTT) – When assessed using the visual occlusion methodology, the NHTSA guidelines provide an implicit upper limit of 24 seconds of total task time. While originally intended for visual/manual tasks, these guidelines provide a reasonable upper limit for task durations of any type.

Visual demand – The visual workload associated with the performance of a task. This would include the structural interference associated with taking the eyes off the forward roadway as well as the central interference in visual processing that arises from cognitive demand. In this report, we refer to the visual demand associated with performing IVIS tasks with different modes of interaction when the vehicle is in motion.

Visual referent task – A variant of the SuRT task (see above) served as the visual referent task in the current research.

Workload – The aggregate of cognitive, visual, and manual demands on the driver. A motorist's workload reflects a combination of demands from the primary task of driving and any secondary tasks performed by the driver. The terms demand and workload are used interchangeably in this report and we develop separate metrics for cognitive workload and visual workload.

ABBREVIATED TERMS

DRT – Detection Response Task

ISO – International Organization for Standardization

IVIS – In-Vehicle Information System

NHTSA – National Highway Traffic Safety Administration

OEM – Original Equipment Manufacturer

SuRT – Surrogate Reference Task

LIST OF FIGURES

INTRODUCTION

2017 model-year automobiles provide a number of features and functions that allow motorists to perform a variety of secondary tasks unrelated to the primary task of driving. Many of these *In-Vehicle Information Systems* (IVIS) involve complex, multimodal interactions to perform a task. For example, to select a music option a driver might push a button on the steering wheel, issue a voice-based command, view options presented on an LCD display located in the center stack, and then select an option via touch using the touch-screen options on the LCD display. Complex multimodal IVIS interactions such as this may distract motorists from the primary task of driving by diverting the eyes, hands, and/or mind from the roadway (Regan, Hallett, & Gordon, 2011; Regan & Strayer, 2014).

Distraction from IVIS interactions arises from a combination of three sources (Strayer, Watson, & Drews, 2011). Impairments to driving can be caused by a competition for visual information processing—for example, when a driver takes his or her eyes off the road to perform IVIS interactions. Impairments can also come from manual interference, as in cases where drivers take their hands off the steering wheel to perform an IVIS interaction. Finally, cognitive sources of distraction occur when attention is withdrawn from the processing of information necessary for the safe operation of a motor vehicle. These sources of distraction can operate independently, but they are not mutually exclusive, and therefore different IVIS interactions can result in impairments from one or more of these sources. In fact, few if any tasks are "process pure" (Jacoby, 1991) and instead often place demands on multiple resources (Wickens, 2008).

Prior research has evaluated workload when motorists performed activities unrelated to driving. For example, the Crash Avoidance Metrics Partnership (CAMP; Angell et al., 2006) investigated the effects of 22 different secondary tasks requiring a combination of visual, manual and cognitive resources on driving performance. Some of the visual-manual tasks required participants to tune the radio or adjust fan speed using physical buttons located in the center console. Auditory-vocal tasks required drivers to listen to an audiobook or sports broadcasts and answer related questions. Distinctive driver-performance profiles suggested that task-induced driver workload was multimodal and characterized by different combinations of visual, manual and cognitive components. In particular, relative to a baseline driving condition, visual-manual tasks were associated with reduced driving-related event detection and more time spent glancing away from the forward roadway. By contrast, auditory-vocal tasks tended to focus drivers' gaze on the forward roadway and resulted in better lane position maintenance – a phenomenon referred to as cognitive tunneling (see, Victor et al., 2005).

In a series of studies, Reimer, Mehler and colleagues (Mehler et al., 2015; Reimer et al., 2014; McWilliams et al., 2015) tested real-world infotainment systems. In Mehler et al. (2015), participants drove two vehicles (2013 Chevrolet Equinox, 2013 Volvo XC60) and interacted with the infotainment systems (MyLink and Sensus, respectively). A combination of ocular measures, subjective workload ratings, and behavioral metrics (e.g., task completion time) was adopted to examine levels of driver workload associated with completing contact calling and navigation-related tasks. Results showed that using visual-manual systems required longer and more frequent off-road glances than auditory-vocal systems. Self-report measures of workload for voice interfaces were higher than those for visual-manual systems. However, the task completion time data showed mixed results, with benefits of auditory-vocal systems observed with MyLink disappearing when drivers used the Sensus system.

Our prior research provided a comprehensive assessment of cognitive workload associated with voice-based interactions, which are known to divert attention from the driving task and lead to cognitive distraction (Strayer et al., 2015, 2016, 2017). We used converging methods to provide a systematic analysis of the workload associated with different voice-based interactions. This included collecting a variety of performance measures (e.g., primary-task measures, secondary-task measures, subjective measures, and physiological measures) to provide a fine-grained assessment of variations in driver workload as they performed different IVIS interactions (e.g., calling and dialing, audio entertainment, text messaging). In Strayer et al., (2016), 257 subjects participated in a weeklong evaluation of the IVIS interaction in one of 10 different model-year 2015 automobiles. After an initial assessment of the cognitive workload associated with IVIS interactions, participants took the vehicle home for five days and practiced using the system. At the end of the five days of practice, they returned and the workload of these IVIS interactions was reassessed. The cognitive workload was found to be moderate to high and was associated with the intuitiveness and complexity of the system and the time it took participants to complete the interaction. Importantly, practice did not eliminate the interference from IVIS interactions. In fact, IVIS interactions that were difficult on the first day were still relatively difficult to perform after a week of practice. There were also long-lasting residual costs after the IVIS interactions had terminated. We suggested that the higher levels of workload should serve as a caution because these voice-based interactions can be cognitively demanding and ought not to be used indiscriminately while operating a motor vehicle.

Task duration is central to the issue of workload assessment. A simple but elegant argument for the importance of task duration has been outlined by Shutko and Tijerina (2006). They suggest that evaluation of task duration is critical not because it reflects a cumulative effect of load but because it represents the time over which an unexpected event might occur. Using a simple exposure based model, they argue that all else being equal, a task that takes twice as long to complete will result in twice the potential risk of an adverse event. [2]

Task duration is commonly measured independently as a stand-alone performance measure or implicitly as compound measure (e.g., Reimer et al., 2014, Ito et al., 2001). Examples of compound duration-related measures include eyes-off-road time, single-glance duration, and total task time (SAE J2944). Formally, duration related measures can be defined as measures that covary with task duration (Burns et al., 2010). Abstractly, duration related measures are those that involve the accumulation of a measured value over time. They can include any measurable performance characteristic related to the vehicle control, secondary task performance, driver behavior, attitudes, thoughts, etc. A key characteristic of duration-based measures is that they are correlated with total task time and change in value with longer tasks (e.g., longer visual tasks result in greater total eyes-off-road time).

Conversely, momentary performance measures provide a summary measure, irrespective of task duration, which characterizes an average slice of task performance. Generally speaking, the stability of momentary performance measures increases with additional sampling. Longer tasks provide more opportunity to measure performance and thus often result in a more stable measure. In driving, common momentary performance measures relate to response time, movement time, lateral control, longitudinal control, steering control, etc. (See SAE J2944 for a comprehensive list). However, momentary performance measures are insufficient to capture variability arising from task duration.

There is no clear consensus on what constitutes an acceptable interaction time for a secondary task. Problematically, the issue is confounded by research suggesting that secondary tasks are often sensitive to whether testing is completed in a static (i.e., not driving) or dynamic (i.e., driving)

[2] Other models suggest a cascading negative effect of task duration on situation awareness (e.g., Fisher & Strayer, 2014).

environment (Young et al., 2005), the age of participants (McWilliams et al., 2015b), and performance characteristics of the primary or secondary tasks (Tsimhoni, Yoo, & Green, 1999). Because of the visual demands associated with driving, visual secondary tasks generally take longer to complete when performed concurrently with driving. Additionally, due to natural aging processes, older adults generally take longer to perform tasks than younger adults. These issues aside, a number of organizations have provided guidance on what constitutes an acceptable secondary task duration (e.g., Japan Automobile Manufacturers Association, 2004; Driver Focus-Telematics Working Group, 2006; NHTSA 2013).

More recently, NHTSA (2013) has issued a set of voluntary guidelines for visual/manual tasks that suggest that tasks should require no more than 12 seconds of Total Eyes Off Road Time (TEORT) to complete. This 12-second rule is based on the societally acceptable risk associated with tuning an analog in-car radio. Using the visual occlusion method, one of two suitable testing approaches specified by NHTSA, motorists are able to interact with in-vehicle technologies while vision is periodically occluded. The testing procedure specifies that vision be occluded in 1.5-second on/off intervals. Tasks that take longer than 24 seconds to complete would thus exceed the testing criteria of 12 seconds of Total Shutter Open Time. While not part of the evaluation criteria, the visual occlusion testing procedure results in a maximum total task time of 24 seconds (i.e., 12 seconds of shutter open time + 12 seconds of shutter closed time for a total of 24 seconds). Despite it being derived from the occlusion method test procedure, we feel that 24 seconds of continuous task engagement represents a reasonable upper limit for task durations of any type.

An important prerequisite for duration-based measures of secondary task performance is the definition of a task. We use the definition provided by Burns et al., (2010) which is a derived from the Alliance of Automobile Manufacturers, ISO, and JAMA Guidelines. Burns et al., suggest that a task can be defined as a sequence of inputs leading to a goal at which the driver will normally persist until the goal is reached. However, we differentiate between continuous and discrete tasks that are shaped by different performance goals. Fundamental to secondary discrete tasks is a performance goal with a finite beginning and end state (e.g., changing the audio source, dialing a phone number, calling a contact, entering a destination into a navigation unit, etc.). Conversely, continuous tasks are characterized by performance maintenance over an indefinite period of time, often with no clear termination state (Schmidt & Lee, 2005) (e.g., conversing via a cell phone, listening to music, following route guidance, etc.). Given the nature of discrete tasks, a failure to account for task duration during assessment provides an incomplete picture of distraction potential.

Research Questions

An important knowledge gap concerns the workload associated with making complex multimodal IVIS interactions. What are the visual and cognitive demands associated with different modes of IVIS interactions (e.g., auditory/vocal interactions versus visual/manual interactions)? To what degree do the different IVIS task types (e.g., audio entertainment, calling and dialing, text messaging, navigation, etc.) place differential demands on visual and cognitive resources? Vehicles clearly differ in their configuration and layout, but do they differ in the visual and cognitive demands of IVIS interactions? Are there trade-offs for IVIS interactions performed with one task or mode of interaction versus another? For example, auditory/vocal inputs may have lower levels of visual demand than issuing commands using a visual/manual touch screen, but the time taken to perform the interaction may be longer in the former than the latter. Surprisingly little is known about how these complex multimodal IVIS interactions impact drivers' workloads. Given the ubiquity of these systems, the current research sought to address three interrelated questions concerning this knowledge gap.

First, *are some task types more impairing than others?* The IVIS interactions support a variety of secondary tasks that are unrelated to the primary task of driving. Some of these interactions may be considered to be so sufficiently impairing that they are locked out by the automaker when the vehicle is in motion (e.g., social media interactions are locked out by most automakers). However, not all secondary tasks are equivalent in distraction potential (e.g., Strayer et al., 2015). They differ in terms of task goals (e.g., play a song, send a text, place a call, etc.). Tasks differ in duration, ranging from a few seconds to a few minutes to complete, with greater distraction potential associated with greater task duration (e.g., Burns et al., 2010). Tasks differ in the way that they are implemented and they may be performed using different modes of interaction (i.e., tasks may be easier to perform using one mode of interaction than another). Tasks may also be performed using a streamlined "one-shot" interaction, or via a series of interactive steps. The current research assessed which task types were most distracting and the degree to which there was any evidence of an interaction between task types and modes of interaction. It is also possible that some tasks may be too demanding to be enabled when the vehicle is in motion, regardless of the mode of interaction.

Second, *are some modes of interaction more distracting than others?* In many instances a task can be performed using auditory/vocal commands, visual/manual interactions, or, as in the example discussed above, a hybrid combination of both auditory/vocal and visual/manual interactions. If the workload associated with one mode of interaction differs from another, the differences may be offset by the time it takes to perform the interaction. For example, a visual/manual touch-screen interaction may divert the driver's eyes from the roadway while an auditory/vocal interaction may keep the eyes on the road; however, if the time to perform an auditory/vocal interaction takes longer than the visual/manual interaction, any benefits of the former may not be realized. Moreover, just because auditory/vocal interactions tend to keep the eyes on the road does not provide a guarantee that drivers will see what they are looking at (Strayer, Drews, & Johnston, 2003, Strayer & Fisher, 2016). The current research is designed to provide an objective benchmark for the level of distraction caused by different modes of IVIS interaction.

Third, *are IVIS interactions easier to perform in some vehicles than others?* A trip to the dealer's showroom will quickly illustrate that vehicles differ in the features, functions, and type of human-machine interface of the IVIS. Are these differences in the IVIS merely cosmetic, or do the differences result in differential workload to perform the same IVIS functions? Vehicles differ in the number and complexity of button interactions on the steering wheel; the size, resolution, and functions supported on center stack LCD display; manual buttons on the center stack and their configuration; and the other unique modes of interaction (e.g., heads-up displays, gesture controls, rotary dials, writing pads, etc.). Moreover, vehicles often provide more than one way to perform a task. There are often cross-modal interactions wherein the task is initiated using one mode of interaction (e.g., voice commands) and then transitions to another mode of interaction (e.g., touch-screen interactions). Some IVIS interactions are ubiquitous (e.g., calling and dialing and audio entertainment), whereas others are supported by one automaker but not another (e.g., destination entry for a navigation system while the vehicle is in motion). The current research compared the IVIS interactions supported by different automakers to determine if they differ in the workload associated with their use. If there are differences in the overall demand of the IVIS interactions, what are the bases for the differences?

Experimental Overview

Our prior research found that it was necessary for the driver to be driving the vehicle in order to accurately assess the concurrent workload associated with IVIS interactions – that is, dynamic testing rather than static testing (cf., SAE J2365, 2016). This was true for IVIS interactions with high levels of cognitive demand, such as using voice commands to interact with the IVIS (e.g., Strayer et al., 2015, 2016, 2017). With cognitive demand, the task of driving added a constant increase to the estimates of driver workload (e.g., the time to perform a purely voice-based IVIS interaction in a moving vehicle was increased by a constant from the time to perform the same interaction in a stationary vehicle). This problem was exacerbated for IVIS interactions with high levels of visual demand, such as making selections on a center stack touch screen, where the time to perform an IVIS interaction in a moving vehicle was an increasing linear function of the time to perform the same interaction in a stationary vehicle. Consequently, all estimates of driver workload in the current research were obtained when participants were driving the vehicle and engaged in IVIS interactions or driving in one of control conditions (i.e., a dynamic testing method). The driving route we used was a low-density residential section of roadway with a speed limit of 25 mph.[3]

To properly scale the driver's workload while interacting with the IVIS, several control conditions were required. First, a single-task driving baseline was needed to estimate the workload of the driver when they were driving the vehicle without the additional workload imposed by the IVIS interactions. This single-task baseline controls for any differences between participants and the workload associated with driving the different vehicles. The single-task baseline anchors the low end of the cognitive and visual workload estimates derived in our research.

To scale cognitive demand, a high workload cognitive task was selected that could be performed in the same way by all participants in all vehicles. The high workload referent task we used was an N-back task (e.g., Mehler, Reimer, & Dusek, 2011; Zhang et al., 2015) in which a prerecorded series of numbers ranging from 0 to 9 were presented at a rate of one digit every 2.5 seconds. Participants were instructed to say out loud the number that was presented two trials earlier in the sequence. The N-back task places a high level of cognitive demand on the driver without imposing any visual demands.[4] Using the single-task baseline and N-back referent provided a way to standardize the cognitive demand of the different IVIS interactions. That is, after controlling for any differences in workload associated with different vehicles using the single-task baseline, IVIS interactions can be directly compared to the N-back task to provide an objective measure of cognitive demand associated with their performance.

To scale visual demand of the IVIS interactions, a high workload visual referent task was selected that could be performed in the same way by all participants in all vehicles. The high workload task we used was a variant of the ISO TS 14198 Surrogate Reference Task (SuRT; Engström & Markkula, 2007; Mattes, Föhl, & Schindhelm, 2007, Zhang et al., 2015) that required participants to

---

[3] This driving route provides a consistent and somewhat conservative estimate of the driving demands. More complex and demanding driving conditions would likely yield higher workload estimates.

[4] In pilot testing, the N-back task used in the current research produced similar levels of overall workload to that of the OSPAN task that we have used in prior research to benchmark high cognitive demand (e.g., Strayer et al., 2015). We chose to use the N-back task in the current research because the workload demands were more consistent over the testing interval and also because the N-Back task was easier to administer.

use their finger to touch the location of target items (larger circles) presented in a field of distractors (smaller circles) on an iPad Mini tablet computer that was mounted in a similar position in all the vehicles. Immediately after touching the location of the target, a new display was presented with a different configuration of targets and distractors. Drivers were instructed to perform the SuRT task while giving the driving task highest priority. The SuRT task places a high level of visual demand on the driver, who must look at the display in order to locate the targets. Using the single-task baseline and SuRT referent provides a way to standardize the visual demand of the different IVIS interactions. That is, after controlling for any differences in workload associated with different vehicles using the single-task baseline, IVIS interactions can be directly compared to the SuRT task to provide an objective measure of visual demand associated with its performance. [5]

The current research used converging performance measures to benchmark the workload of the IVIS interactions. This included the collection of *subjective* estimates from the driver on their workload using the NASA-Task Load Index (Hart & Staveland, 1988) at the end of testing each IVIS interaction.

We also assessed driver workload using the Detection Response Task (DRT), an International Standards Organization (ISO) protocol for measuring attentional effects of cognitive load in driving (ISO 17488, 2015). The DRT procedure involves presenting a secondary stimulus (e.g., a changing light or vibrating buzzer) every three to five seconds and requiring the driver to respond to these events when they detect them by pressing a microswitch attached to their thumb to the steering wheel.[6] As the workload of driving and/or the IVIS interactions increase, the reaction time to the DRT stimulus increases and the likelihood of detection of the DRT stimulus (i.e., the hit rate) decreases (e.g., Strayer

[5] The N-back referent induces a high level of cognitive demand and the task does not present any visual information for the driver to look at. However, it is well known that high levels of cognitive demand often alter the visual scanning behavior of the driver (e.g., see Strayer & Fisher, 2016 for a review). That is, the N-back task may impair what the driver sees. Similarly, the SuRT referent induces a high level of visual demand by requiring the driver to look at a touch screen to locate a target amongst distractors. However, in addition to taking driver's eyes off the roadway to perform the task, visual attention is required to perform the SuRT task (i.e., the SuRT task we used was a feature search task e.g., Triesman & Gelade, 1980). Pilot testing of the SuRT task found a visual search slope of approximately 20 msec/item, a value above the upper threshold associated with automatic visual search (e.g., Schneider & Shiffrin, 1977; Shiffrin and Schneider, 1977). Thus, the SuRT task has high visual/manual demand and modest cognitive demand.

[6] Despite the fact that the processing requirements of the DRT are minimal, it is possible that its inclusion could increase the workload of the driver compared to conditions without the DRT test. In our earlier work, we compared the subjective workload of one group of drivers who were using the DRT with that of another group who performed the same tasks without the DRT (see Strayer et al., 2013) and found that the DRT did not increase the workload of the driver. Using a within-subjects design, Castro, Strayer, Matzke, & Heathcote (2017) compared pursuit-tracking performance with and without the DRT and found minimal changes in tracking performance between the two tracking conditions (i.e., the inclusion of the DRT did not impair primary-task tracking performance). A similar analysis using the DRT to assess workload in a complex multitasking study, Palada, et al. (2017) also found that the DRT had minimal (and non-significant) effects on primary task performance (i.e., < 10 msec). On the whole, there is scant evidence that inclusion of the DRT significantly alters performance of the driving task.

et al., 2015, 2016, 2017). The DRT has proven to be very sensitive to dynamic changes in drivers' workload (e.g., Strayer, Biondi, & Cooper, 2017). The DRT provides an *objective* assessment of the driver's workload associated with different IVIS interactions.

We used two variants of the DRT in our research. The first variant was a vibrotactile DRT, in which a vibrating buzzer, which feels similar to a vibrating smartphone, was attached to the participant's left collarbone and a microswitch was attached to a finger on the driver's left hand so that it could be depressed against the steering wheel when they detected the vibration. The vibrotactile DRT provides a sensitive measure of the participant's cognitive load as they perform different IVIS interactions. As the cognitive demand increases, the RT to the vibrotactile DRT increases. These RT differences were calibrated by using the single-task baseline and N-back referent to anchor the workload of the IVIS interactions in different vehicles.

Specifically, evaluation of the cognitive demand of any IVIS interaction involved an initial subtraction from any differences between vehicles and/or participants obtained in the single-task baseline (i.e., this defined the relative demand associated with an IVIS interaction). This relative cognitive demand was compared to the N-back task (i.e., the difference between the N-back task and single-task baseline defined the relative cognitive demand of the N-back task). The Cognitive Demand Ratio (CDR) was defined as the ratio of the relative cognitive demand of an IVIS interaction to the relative cognitive demand associated with the N-back task.

The CDR provides a standardized metric for comparison across IVIS interactions (both within a vehicle and between vehicles). For example, if an IVIS interaction has a CDR that is between 0 and 1, the cognitive demand of that interaction is greater than the single-task baseline and less than the N-back task. If an IVIS interaction has a CDR greater than 1, then the cognitive demand of that IVIS interaction exceeds the N-back task. Furthermore, if the CDR of an IVIS interaction in one vehicle is greater than the same IVIS interaction in another vehicle, the two vehicles differ in the cognitive demand of that interaction, with the former being greater than the latter.

The second variant of the DRT used a light that was projected onto the windshield in the driver's line of sight as they looked at the forward roadway. When the DRT light changed from orange to red, the participant was instructed to press the microswitch attached to their finger when they detected the changing light (the same response that was used for the vibrotactile DRT). The remote DRT provides a sensitive measure of the participant's visual load as they perform different IVIS interactions. As the visual demand increases, the detection of the changing light decreases (i.e., a decrease in hit rate). These hit rate differences were calibrated using the single-task baseline and SuRT task to anchor the workload of the IVIS interactions in different vehicles.

Evaluation of the visual demand of any IVIS interaction involved an initial subtraction from any differences between vehicles and/or participants obtained in the single-task baseline (i.e., this defined the relative visual demand associated with an IVIS interaction). This relative visual demand was compared to the SuRT task (i.e., the difference between the SuRT referent and single-task baseline defined the relative visual demand of the SuRT task). The Visual Demand Ratio (VDR) was defined as the ratio of the relative visual demand of an IVIS interaction to the relative visual demand associated with the SuRT task.

As with CDR, VDR provides a standardized metric for comparison across IVIS interactions (both within a vehicle and between vehicles). For example, if an IVIS interaction has a VDR that is between 0 and 1, the visual demand of that interaction is greater than the single-task baseline and less than the SuRT task. If an IVIS interaction has a VDR greater than 1, then the visual demand of that

IVIS interaction exceeds the SuRT task. Furthermore, if the VDR of an IVIS interaction in one vehicle is greater than the same IVIS interaction in another vehicle, the two vehicles differ in the visual demand of that interaction, with the former being greater than the latter.[7]

In order to capture the effects of task duration, our measures of momentary cognitive, visual, and subjective task demand were scaled by task completion time. Tasks that took longer than 24 seconds resulted in an upward bias of overall demand, whereas tasks that took less than 24 seconds resulted in a downward bias. Of the metrics that fed into the overall workload metric, total task time is most amenable to modification through design. Our investigation found that factors such as menu depth, display clutter, system responsivity, dialog verbosity, cellular connection stability, and server performance all play a significant role in task duration (e.g., Biondi, Getty, Cooper, & Strayer, under review; Getty, et al., under review). The time required for a user to complete a task can be reduced through the careful performance evaluation, resulting in a reduction in exposure duration.

A total of 24 participants drove each of the vehicles tested in the experiment. The duration of a testing session for a vehicle was dependent on the features and functions available in each vehicle (testing ranged from two-and-a-half to three-and-a-half hours). The single-task and two referent tasks and the IVIS interactions tested in each vehicle were evaluated in an experimental order that was counterbalanced across participants. The evaluation of the IVIS interactions facilitated a comparison of different task types (e.g., calling and dialing, audio entertainment, navigation, and text messaging), modes of interaction (e.g., auditory vocal, center stack, center console), and vehicles. From this design, it was possible to determine the effects of cognitive and visual demand associated with different task types, modes of interactions, vehicles, and the interaction of each of these factors.

METHOD

Participants

[7] The remote DRT is sensitive to where drivers are looking. The logic of the measure is straightforward. When an IVIS interaction requires participants to divert their eyes from the forward roadway, they are less likely to detect the changing light. Because pilot testing found that onset visual cues could be detected even when participants were looking away from the forward roadway, we made the visual "off" stimulus an orange light and the visual "on" stimulus a red light (i.e., the task required the detection of the light changing from orange to red). Pilot testing found that drivers could easily detect the light change when they were looking at the forward roadway. To make sure that the remote light stimulus was sensitive to visual demand (as opposed to cognitive demand), we performed a series of validation studies in a driving simulator using an eye tracker that indicated that the pattern of hit rates using the remote DRT matched the eye tracking estimates of eyes off the road (for details, see Castro, Cooper, & Strayer, 2016; Cooper, Castro, & Strayer, 2016). Note that because of varying lighting conditions, eye trackers often prove difficult to use in on-road testing situations.

The hit rate to the remote DRT is a measure of drivers seeing what they are looking at as they direct their gaze at the forward roadway (and similarly, miss rates to the remote DRT are evidence of drivers not looking at and/or not seeing information on the forward roadway). However, cognitive demand can also impair detection rates to the remote DRT, as documented by the inattention blindness literature (e.g., Strayer, Drews, & Johnston, 2004).

After approval from the University of Utah IRB, 120 participants (54 female) with an age range of 21-36 years (M = 25) and a reported average of 9.1 driving hours per week were recruited via flyers and social media. All participants were native English speakers, had normal or corrected-to-normal vision, held a valid driver's license and proof of car insurance, and had not been the at-fault driver in an accident within the past two years. Compensation was prorated at $20 per hour. Prior to participation, a Motor Vehicle Record report was obtained by the University of Utah's Division of Risk Management to ensure a clean driving history. Each participant was also required to complete a 20-minute online defensive driving course and pass the accompanying certification test, as per University of Utah policy.

Twenty-four participants were tested in each vehicle. We used a planned missing data design (e.g., Graham, Taylor, Olchowski, & Cumsille, 2006; Little & Rhemtulla, 2013) where some of the participants were tested in just one of the vehicles, whereas other participants were tested on multiple vehicles on separate occasions. On average, participants were tested on five vehicles, with a range of one to 24 vehicles (e.g., one participant was tested in 24 of the vehicles).[8] Participants were initially unfamiliar with the specific IVIS tasks and systems but were trained until they felt comfortable performing each of the requested interactions. Additionally, participants gained broad experience with the different systems, tasks, and modes of interaction offered by each vehicle through repeated research participation.

Stimuli and Apparatus

The vehicles used in the study are listed below. Vehicles were selected for inclusion in the study based on an initial assessment of market share of the vehicle, the IVIS features available in the vehicle, and availability of vehicles for testing. This sample was representative of 30% of the market share in North America. Vehicles were acquired through Enterprise Rent-A-Car and short-term leases from automotive dealerships or purchased for testing. A complete description of the features and functions available in each vehicle is provided in Table 1. Obviously, the specific sequence of commands to perform the different tasks varied as a function of OEM and modality of interaction. The specific syntax and command sequence to perform the different tasks in each of the vehicles and modalities of interaction are provided in Appendix A.

- 2017 Audi Q7 3.0T Premium Plus
- 2017 Cadillac XT5 Luxury
- 2017 Chevrolet Equinox LT
- 2017 Chevrolet Traverse LT
- 2017 Chrysler 300C
- 2017 Dodge Durango GT
- 2017 Dodge Ram 1500 Express
- 2017 Ford F-250 XLT

[8] The number of vehicles driven by a participant was associated with the overall demand score ($b = -0.02$, $t = -3.38$, $p = <.001$). However, the effect size of number of vehicles driven was relatively small, accounting for ~10% of the variability between participants. Though modest, we retained the number of vehicles driven by participants in all linear mixed effects models presented below in order to control for any impact of this variable.

- 2017 Ford Fusion Titanium
- 2017 Ford Mustang GT Premium Convertible
- 2017 GMC Yukon SLT
- 2017 Honda Civic Touring
- 2017 Honda Ridgeline RTL-E
- 2017 Hyundai Santa Fe Sport
- 2017 Hyundai Sonata Base
- 2017 Infiniti Q50 3.0t Premium
- 2017 Jeep Compass Sport
- 2017 Jeep Grand Cherokee Limited
- 2017 Kia Sorento LX
- 2017 Lincoln MKC Premiere
- 2017 Mazda3 Touring
- 2017 Nissan Armada SV
- 2017 Nissan Maxima SV
- 2017 Subaru Crosstrek Premium
- 2017 Tesla Model S 75
- 2017 Toyota Camry SE
- 2017 Toyota Corolla SE
- 2017 Toyota RAV4 XLE
- 2017 Toyota Sienna XLE
- 2017 Volvo XC60 T5 Inscription

Equipment

Identical LG K7 android phones on the T-Mobile mobile network were paired via Bluetooth with each vehicle. Each vehicle was also equipped with two Garmin Virb XE action cameras, one mounted under the rear-view mirror to provide recordings of participants' faces, and an additional camera mounted near the passenger seat shoulder to provide a view of the dash area for infotainment interaction. Video was recorded at 30 frames per second, at 720p resolution. An iPad Mini 4 (20.1 cm diagonal LED-backlit Multi-Touch display) was connected to each vehicle via USB and was pre-loaded with a small music library. Identical Acer R11 laptop computers were utilized for data collection in the vehicle.

Stimuli

Participants completed tasks requiring IVIS interaction. Depending on the vehicle, participants would interact with the system to perform tasks involving audio entertainment, calling and/or dialing, navigation, and text messaging. Also dependent on vehicle interface was the method by which participants would interact (see Appendix for complete details of each vehicle). All vehicles had voice recognition; however, vehicles differed on visual/manual interaction (e.g., touch screen, manual buttons, rotary wheel, and wheel pad). The interaction tasks in each vehicle were matched as closely as possible given the differences in the systems' capabilities.

DRT

Participants were required to respond to a vibrotactile and remote DRT as per ISO 17488 (2015). A vibrotactile device was placed on the participant's left collarbone area and a microswitch was attached to either the index or middle finger of the left hand so that it could be depressed against the steering wheel. A remote DRT light was placed along a strip of Velcro on the dashboard in such a way that the participant could not directly gaze upon the light but instead saw the reflection in the windshield directly in their line of sight (Castro, Cooper, & Strayer 2016; Cooper, Castro, & Strayer, 2016). Millisecond resolution response time to the vibrotactile onset or LED light was recorded via an embedded micro-controller and stored on the host computer.

Following the ISO guidelines (2015), the vibrotactile device emitted a small vibration stimulus, similar to a vibrating cell phone. The LED light stimulus was a change in color from orange to red. These changes cued the participant to respond as quickly as possible by pressing the microswitch against the steering wheel. The tactor and light were equiprobable and programmed to occur every three to five seconds (with a rectangular distribution of inter-stimulus intervals within that range). They lasted for one second or until the participant pressed the microswitch. The task of driving was considered the primary task, the interaction with the IVIS was considered the secondary task, and responding to the DRT was considered a tertiary task.

Procedure

Participants completed tasks involving interacting with the infotainment system in the vehicle to achieve a particular goal (i.e., using the touch screen to tune the radio to a particular station, using voice recognition to find a particular navigation destination, etc.) while driving. Tasks were categorized into one of four task types: audio entertainment, calling and dialing, text messaging, and navigation, depending on vehicle capabilities. These task types were completed via different modalities equipped in each vehicle (i.e. touch screen, voice recognition, rotary wheel, draw pad, etc.) for each interaction. The order of interactions was counterbalanced across participants.

The possible task types performed by the participant are listed below. A description of the specific tasks performed in each vehicle is provided in Table 2.[9]

- *Audio Entertainment*: Participants changed the music to different FM and AM stations, a satellite radio source, the LG K7 phone connected via Bluetooth, and the Mini iPad connected via USB.

- *Calling and Dialing*: A list of 91 contacts with a mobile and/or work number was created for participant use. In vehicles capable of dialing phone numbers, participants were instructed to dial the phone number 801-555-1234 as well as their own phone number.

[9] A master reset of the IVIS system was performed before testing commenced.

- *Text Messaging*: Depending on the texting capabilities of each vehicle, participants either listened to short text messages sent by other LG K7 phones or sent a new text from the list of predetermined messages specific to each vehicle.

- *Navigation*: Participants started and canceled route guidance to different local and national businesses that differed according to the options presented by each system.

The potential modes of interaction performed by the participant are listed below. A description of the interaction modes for each vehicle is provided in Table 2. Interaction modalities were selected and individual tasks created based on vehicle capabilities.

- *Center Stack*: The center stack is located in the center of the dash to the right of the driver. A visual display is used to present textual and/or graphical information. Center stack systems often include a touch-screen interface to support visual/manual interactions so that drivers can select an option and navigate menus by touch and/or use slider bars to scroll through options displayed on the screen. With some vehicles, the selection of options may be made with manual buttons surrounding the touch screen.

- *Auditory Vocal*: A voice-based interaction is initiated by the press of a physical button on the steering wheel or center stack. Microphones installed in the vehicle pick up the driver's voice commands and process them to perform specific functions and access help menus in the vehicle. Possible voice command options may be presented aurally or displayed on the vehicle's center stack to aid the driver in making valid commands.

- *Center Console*: The center console is located between the driver and passenger front seats. The interactions are made through a rotary dial that allows drivers to scroll through menu items presented on the center stack visual display. Another interaction variant uses a writing pad where drivers used their finger to write out commands.

Driving Route

A low-traffic residential road with a 25 mph speed limit was used for the on-road assessment. The route contained four stop signs and two speed bumps. The participants were required to follow all traffic laws and adhere to the 25 mph speed limit at all times. The length of road was approximately two miles one way, with an average drive time of six minutes in each direction (see Figure 1). A researcher was present in the passenger seat of each vehicle for safety monitoring and data collection.

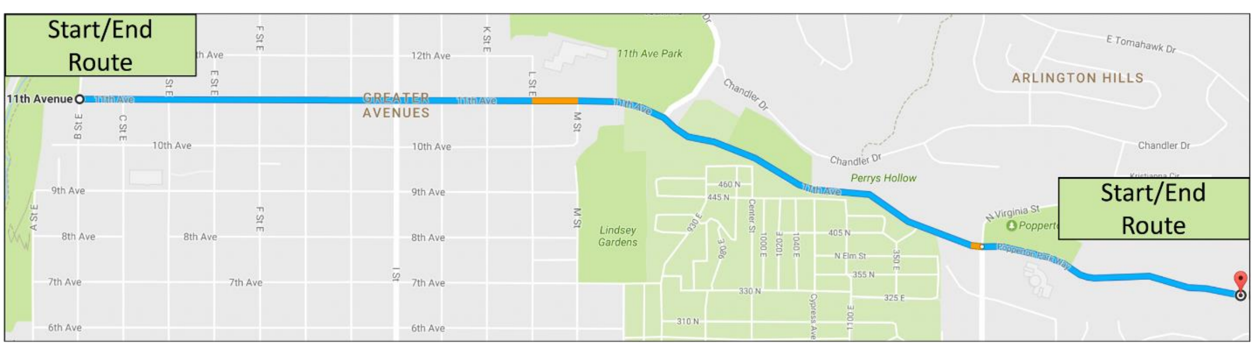

Figure 1. A bird's-eye view of the driving route used in the study.

Training

Before the study commenced, participants were given time to adjust and familiarize themselves with the vehicle while driving a practice run on the designated route. During the familiarization drive, the researcher pointed out potential road hazards. After participants felt comfortable in the vehicle, they were trained on how to respond to the DRT. The researcher verified that participants responded appropriately to 10 stimuli presented between three to five seconds apart and that they had response times of less than 500 milliseconds. Next, they were trained on how to interact with and complete tasks via a particular modality before each condition began. In order to be considered properly trained, participants were required to perform three trials without error immediately before the testing commenced for each of the IVIS interactions. Once participants expressed confidence in their ability to interact with the system, the experimental run began.

Participants were instructed to drive the designated route from one end to another, performing the IVIS interactions as instructed by the experimenter several times on each drive. When the participant reached the end of the route, they were instructed to pull over, marking the end of one of the experimental blocks. The next experimental block began in the opposite direction on the designated route, and this process was repeated until all conditions had been completed.

While driving, verbal task instructions provided by the researcher were given to the participant (i.e., "using the touch screen, tune radio to 96.3 FM"). Participants were instructed to not initiate the task until the researcher told them to do so by saying, "go." Once the given task was complete, the participant would say, "done." The researcher would mark the task start and end time of each task by depressing a key on the data collection computer for later association with timing of on-task performance. DRT trials were considered valid for statistical analysis if they fell between these start and end times. Participants were allowed to take as much time as needed to complete each task. A minimum 10-second interval was provided between tasks. The total number of tasks in each 2-mile run varied with task duration, ranging from five to 11. Participants also performed three control tasks while driving the designated route. The control tasks were:

- *Single-task Baseline*: Participants performed a single-task baseline drive using the vehicle being tested on the designated route without interacting with the IVIS. During the single-task baseline, participants responded to the DRT stimuli.

- *Auditory N-back task*: The auditory N-back task presented a pre-recorded series of numbers at a rate of one digit every 2.5 seconds. Participants listened to auditory lists of numbers ranging from 0 to 9 presented in a randomized order. They were instructed to say out loud the number that was presented two trials earlier in the sequence. Participants were instructed to respond as accurately as possible to the N-back stimuli and the research assistant monitored performance in real-time. During the auditory N-back task, participants also responded to the DRT stimuli.

- *SuRT task*: The SuRT task[10] presented a target on the display with 21 to 27 distractors. The target was an open circle 1.5 cm in diameter and the distractors were open circles 1.2 cm in diameter. The SuRT task, illustrated in Figure 2, was presented on an iPad Mini 4 with circles printed in black on a white background. The participant's task was to touch the location of the target. Immediately thereafter, a new display was presented with a different configuration of targets and distractors. The location of targets and distractors was randomized across the trials in the SuRT task. Participants were instructed to continuously perform the SuRT task while giving the driving task highest priority, and the research assistant monitored performance in real-time. A research assistant instructed participants to pause the SuRT task at intersections or if there were potential hazards on the roadway. During the SuRT task, participants also responded to the DRT stimuli.

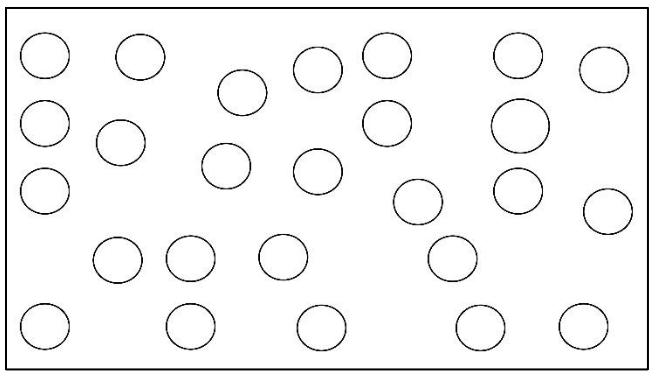

Figure 2. An example of the SuRT task that required participants to touch the location of the target circle.

After the completion of each condition, participants were given a NASA-TLX (Hart & Staveland, 1988) to assess the subjective workload of that car's system.

Dependent Measures

DRT data were cleaned following procedures specified in ISO 17488 (2015). Consistent with the standard, all responses briefer than 100 msec (0.6% of the total trials) or greater than 2,500 msec (1.4% of the total trials) were rejected for calculations of Reaction Time. Nonresponses or responses that occurred later than 2.5 seconds from the stimulus onset were coded as misses. During testing of the IVIS interactions, on-task engagement was recorded by the researcher through a key press on the DRT host computer, which allowed the identification of segments of the IVIS condition when the participant was actively engaged in an activity or had finished that activity and was operating the vehicle without IVIS interactions. Incomplete, interrupted, or otherwise invalid tasks were marked with a key-flag

[10] The variant of the SuRT task we used in the current research matched as closely as possible the visual display characteristics described in ISO/TS 14198; however, participants responded to the target by pressing the touch-screen location rather than using a keypad. This task places visual/manual demands on the driver that are more similar in nature to interactions using the center stack LCD touch screen. Video coding of eye glances when participants performed the SuRT task indicated that they took their eyes off the road for 50% of the time when interacting with the task.

and disregarded from analysis. In addition, task time data were cleaned to remove any recorded tasks with a duration shorter than three seconds that resulted in the removal of less than 0.3%. The dependent measures obtained in the study are listed below:

- DRT – Reaction Time: Defined as the sum of all valid reaction times to the DRT task divided by the number of valid reaction times.
- DRT – Hit Rate: Defined as the number of valid responses divided by the total number of valid stimuli presented during each condition.

Following each drive, participants were asked to fill out a brief questionnaire that posed eight questions related to the just completed task. The first six of these questions were from the NASA TLX; the final two assessed the intuitiveness and complexity of the IVIS interactions.

- Subjective Measures – Defined as the response on a 21-point scale for each question:
  - Mental – How mentally demanding was the task?
  - Physical – How physically demanding was the task?
  - Temporal – How hurried or rushed was the pace of the task?
  - Performance – How successful were you in accomplishing what you were asked to do?
  - Effort – How hard did you have to work to accomplish your level of performance?
  - Frustration – How insecure, discouraged, irritated, stressed, and annoyed were you?
  - Intuitiveness – How intuitive, usable, and easy was it to use the system?
  - Complexity – How complex, difficult, and confusing was it to use the system?

Task Interaction Time was obtained from the time stamp on the DRT host computer. Task interaction time was defined as the time from the moment participants first initiated an action to the time when that action had terminated and the participant said, "done."

Data Analysis and Modeling

The DRT data were used to provide empirical estimates of the cognitive and visual demand of the different conditions. For an estimate of cognitive demand, the average RT to the vibrotactile DRT for each participant was computed for the single-task baseline condition and for the N-back task. Equation 1 was used to standardize the vibrotactile DRT data.

$$Equation\ 1\text{:}\ Cognitive\ Demand = \frac{IVIS\ Task - Single\ Task}{Nback\ Task - Single\ Task}$$

Using Equation 1, the single-task baseline would receive a rating of 0.0 and the N-back task would receive a score of 1.0. IVIS tasks tested in the vehicle were similarly scaled such that values below 1.0 would represent a cognitive demand lower than the N-back task and values greater than 1.0 would denote conditions with a higher cognitive demand than the N-back task. Note that the cognitive demand is a continuous measure ranging from 0 to ∞, with higher values indicating higher levels of cognitive demand.

For an estimate of visual demand, the average hit rate to the remote DRT for each participant was computed for the single-task baseline condition and for the SuRT task. Equation 2 was used to standardize the data collected from the remote DRT.

$$Equation\ 2\colon\ Visual\ Demand = \frac{Single\ Task - IVIS\ Task}{Single\ Task - SuRT\ Task}$$

Using Equation 2, the single-task baseline would receive a rating of 0.0 and the SuRT task would receive a score of 1.0. IVIS tasks tested in the vehicle were similarly scaled such that values below 1.0 would represent visual demand lower than the SuRT task and values greater than 1.0 would denote conditions with visual demand higher than the SuRT task. As with cognitive demand, the visual demand is a continuous measure ranging from 0 to ∞, with higher values indicating higher levels of visual demand.

For an estimate of subjective demand, the average of the six NASA TLX ratings for each participant were computed for the single-task baseline condition and for the N-back and SuRT tasks. Equation 3 was used to standardize the subjective estimates.

$$Equation\ 3\colon\ Subjective\ Demand = \frac{IVIS\ Task -\ Single\ Task}{\left(\frac{Nback\ Task + SuRT\ Task}{2}\right) - Single\ Task}$$

Using Equation 3, the single-task baseline would receive a rating of 0.0 and average of the N-back and SuRT tasks would receive a score of 1.0. IVIS tasks tested in the vehicle were similarly scaled such that values below 1.0 would represent a subjective demand lower than the average of the N-back and SuRT tasks and values greater than 1.0 would denote conditions with subjective demand higher than the average of the N-back and SuRT tasks. As with cognitive demand, the subjective demand is a continuous measure ranging from 0 to ∞, with higher values indicating higher levels of subjective demand.

Equation 4 was used to standardize the IVIS interaction time data using the 24-second interaction time referent (NHTSA, 2013).

$$Equation\ 4\colon\ Interaction\ Time = \frac{IVIS\ Task}{24\ seconds}$$

Using Equation 4, a task interaction time of 24 seconds would receive a score of 1.0. IVIS interactions tested in the vehicle were scaled such that values below 1.0 would represent a task interaction time lower than 24 seconds and values greater than 1.0 would denote conditions with a task interaction time greater than 24 seconds. The time-on-task metric is a continuous measure ranging from 0 to ∞, with higher values indicating longer task interaction time.[11]

[11] The 24-second task interaction referent is derived from NHTSA (2013). Performance on the high visual/manual demand SuRT for 24 seconds, a score of 1.0 in our rating system, matches the threshold for total task time using the Visual Occlusion testing procedure. The general principle is that these multimodal IVIS interactions should be able to be performed in 24 seconds or less when paired with the task of operating a moving motor vehicle.

An overall workload rating was determined by combining the cognitive, visual, and subjective demand with the interaction time rating using Equation 5. Using Equation 5, overall demand is a continuous measure ranging from 0 to ∞, with higher values indicating higher levels of workload.

$$Equation\ 5\colon\ Overall\ Demand = \frac{(Cognitive\ + Visual\ + Subjective)}{3} * Interaction\ Time$$

Application of these formulae provide stable workload ratings with useful performance criteria that are grounded in industry standard tasks. On occasion, however, the approach can return extreme values when either the numerator is unusually small or the task time unusually long. In order to mitigate the potential for unusual scores to skew the overall rating, scores greater than 3.5 standard deviations from the mean (<1% of the data) were excluded from analysis.

Experimental Design

The experimental design was a 4 (Task Type) X 3 (Modality of Interaction) X 30 (Vehicle) factorial with 24 participants evaluated in each vehicle. However, as noted in Table 1, not all vehicles offered the full factorial design (i.e., the Task Type by Modality of Interaction factorial was not always available from all OEMs). Moreover, participants were tested in a varying number of the vehicles. Consequently, a planned missing data design (e.g., Graham, Taylor, Olchowski, & Cumsille, 2006; Little & Rhemtulla, 2013) was used where some cells in the factorial were missing and, as noted above, the number of vehicles driven by a participant was used in all linear mixed effects models presented below in order to control for any impact of this latter factor.

RESULTS

A bootstrapping procedure was used to estimate the 95% confidence intervals (CI) around each point estimate in the analyses reported below. The bootstrapping procedure used random sampling with replacement to provide a nonparametric estimate of the sampling distribution. In our study, there were N=24 participants tested in each vehicle. The bootstrapping procedure involved generating 10,000 bootstrapping samples, each of which were created by sampling with replacement N samples from the original "real" data. From each of the bootstrap samples, the mean was computed and the distribution of these means across the 10,000 samples was used to provide an estimate of the standard error around the observed point estimate.[12]

The greater the spread of the CI, the greater the variability associated with the point estimate. The obtained 95% CI also provides a visual depiction of the statistical relationship between the point estimate and the single-task baseline and/or the high demand referents for cognitive, visual, subjective, and interaction time. For example, if the high demand referent does not fall within the 95% CI, then the point estimate significantly differs from that referent. Similarly, if the 95% confidence intervals of two conditions do not overlap, then the two conditions differ significantly. However, the 95% CI of two conditions may overlap and the differences may still be significant. In this case, if the pair-wise difference between two conditions divided by the pooled standard error exceeds $t(23)=2.064$, the difference is significant at the $p<.05$ level (two tailed).

[12] Prior to bootstrapping, all scores were baseline corrected, minimizing the potential for violations of homogeneity of variance in resampling procedures (e.g., Davidson, Hinkley, & Young, 2003). The baseline correction eliminated any effects of participant in the analyses reported below.

The standardized scores for the high demand cognitive or visual referent tasks can also be translated into effect size estimates (i.e., Cohen's d). For cognitive demand, a standardized score of 1.0 reflects a Cohen's d of 1.423. For visual demand, a standardized score of 1.0 reflects a Cohen's d of 1.519. The high demand estimates for cognitive and visual referent tasks reflect *very large* effect sizes. Note that a standardized score of 2 would reflect a doubling of the effect size estimates, a standardized score of 3 would reflect a tripling of the effect size estimates, and so on. Note also that the effect size estimates for the high cognitive and visual demand are virtually equivalent (differing by less than 0.1 Cohen's d units).

Linear mixed effects analyses were performed using R 3.3.1 (R Core Team, 2016), lme4 (Bates, Maechler, Bolker, & Walker, 2015), and multcomp (Hothorn, Bretz, & Westfall, 2008). In the analyses reported below, Task Type, Modality, Task Type by Modality, and Vehicle were entered independently. Examples of how these contrasts were performed using the lmer function in the lme4 package are presented in Appendix 2. The number of vehicles driven by participant was entered as a fixed effect while Participant, Vehicle, Modality, and Task Type were entered as random effects. In each case, p-values were obtained by likelihood ratio tests comparing the full linear mixed effects model to a partial linear mixed effects model without the effect in question. This linear mixed modeling analysis has the advantage of analyzing all available data while adjusting fixed effect, random effect, and likelihood ratio test estimates for missing data.

In the first section of the results, the data are collapsed over the participants and vehicles to provide an understanding of how workload varied as a function of the task type and mode of IVIS interaction. These analyses are important because they document the demand of the task types and modes of interaction on driver workload independent of vehicle. The next section presents data at the vehicle level, followed by analysis at the vehicle by task type, and vehicle by modes levels of analysis. This section provides a detailed account for *how* and *why* the vehicles differed in workload as participants performed the different IVIS interactions.

Effects of Task Type

Figures 3-7 present the workload associated with the *four IVIS task types* evaluated in the on-road testing. Figure 3 presents the cognitive demand, Figure 4 presents the visual demand, Figure 5 presents the subjective demand, and Figure 6 presents the task interaction time. The overall demand is presented in Figure 7.

Cognitive demand was derived using Equation 1 and inspection of Figure 3 shows that the cognitive demand from each task type was greater than the N-back task (i.e., in each case the cognitive demand exceeded the red vertical line). The relative ordering of task types placed calling and dialing and the navigation task types as slightly less cognitively demanding than the audio entertainment and texting task types. This conclusion was confirmed by a significant difference in fit of a linear mixed effects models with and without task type included ($\chi^2(3) = 11.07$, $p = .011$).

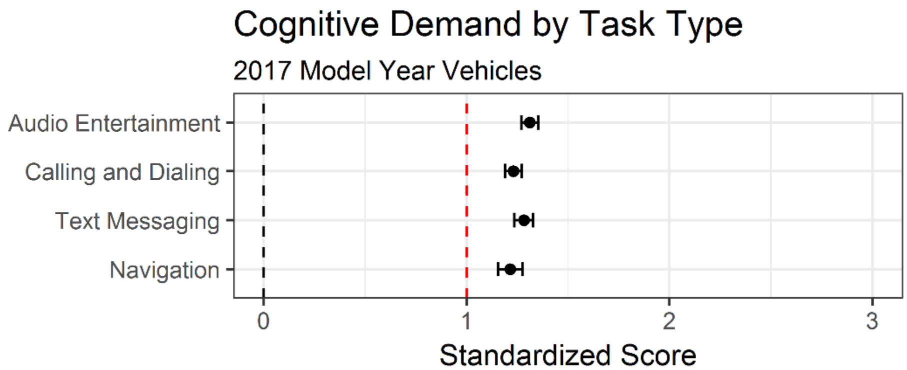

Figure 3. Cognitive demand as a function of task type for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the N-bask task. Error bars represent 95% confidence intervals.

Visual demand was derived using Equation 2. A comparison of linear mixed effects models with and without Task Type indicated that Task Type was a significant predictor of visual demand ($\chi^2(3) = 55.3$, $p < .001$). Inspection of Figure 4 shows that the visual demand was not significantly different from the SuRT task for calling and dialing and text messaging task types but was significantly higher than the SuRT referent for the audio entertainment and navigation task types. The overlap in confidence intervals indicates that the audio entertainment and navigation task types did not significantly differ in visual demand.

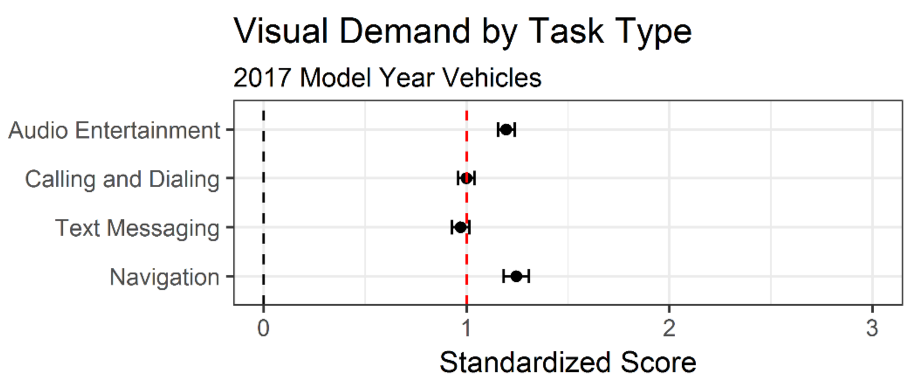

Figure 4. Visual demand as a function of task type for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the SuRT task. Error bars represent 95% confidence intervals.

Subjective demand was derived using Equation 3. A comparison of linear mixed effects models with and without Task Type indicated that Task Type was a significant predictor of subjective demand ($\chi^2(3) = 26.69$, $p < .001$). Inspection of Figure 5 shows that the subjective demand of all of the task types was less than the average of high demand referent tasks. The relative ordering of the task types placed calling and dialing below the audio entertainment, texting and the navigation task types. However, the overlap in confidence intervals indicates that the task types did not significantly differ in subjective demand with the exception of the contrast between calling and dialing and navigation.

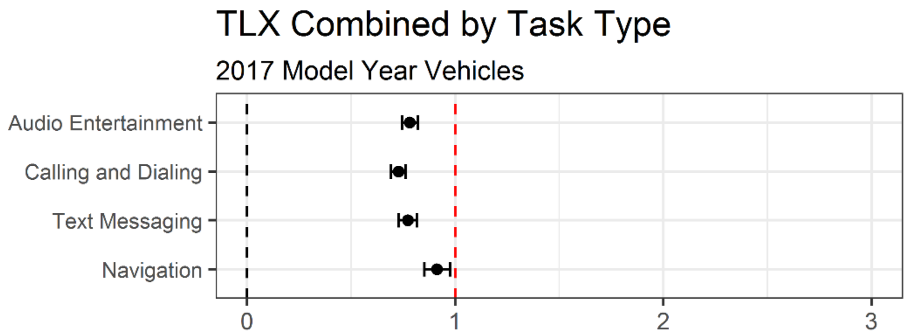

Figure 5. Subjective demand as a function of task type for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the average demand of the N-back and SuRT tasks. Error bars represent 95% confidence intervals.

Task interaction time was derived using Equation 4. A comparison of linear mixed effects models with and without Task Type indicated that Task Type was a significant predictor of interaction time ($\chi^2(3) = 2456$, $p < .001$). Inspection of Figure 6 shows that text messaging and navigation task types took significantly longer than the 24-second interaction referent. The audio entertainment task type took significantly less time than the calling and dialing task type, which took less time to perform than the text-messaging task type. The longest task interaction times were associated with navigation, which took an average of approximately 40 seconds to complete.

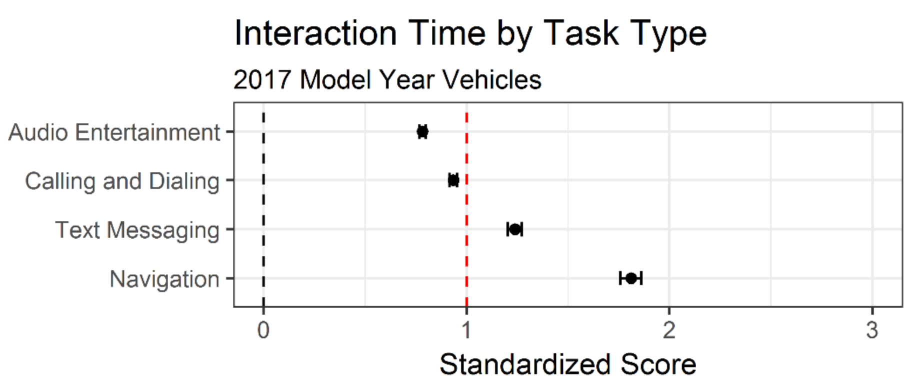

Figure 6. Interaction time as a function of task type for the on-road assessment. The dashed vertical red line represents the 24-second task interaction referent. Error bars represent 95% confidence intervals.

Finally, overall demand, derived using Equation 5 and presented in Figure 7, shows that demand of the audio entertainment and calling and dialing task types fell below the high workload benchmark represented by the red vertical line and the text messaging and navigation task types exceeded the standardized high workload benchmark. A comparison of linear mixed effects models with and without Task Type indicated that Task Type was a significant predictor of overall demand ($\chi^2(3) = 962$, $p < .001$). Of the four task types evaluated, audio entertainment and calling and dialing were the easiest to perform and they did not significantly differ in overall demand. Text messaging was significantly more demanding than audio entertainment and calling and dialing. The navigation task type was significantly more demanding than any of the other task types that were evaluated.

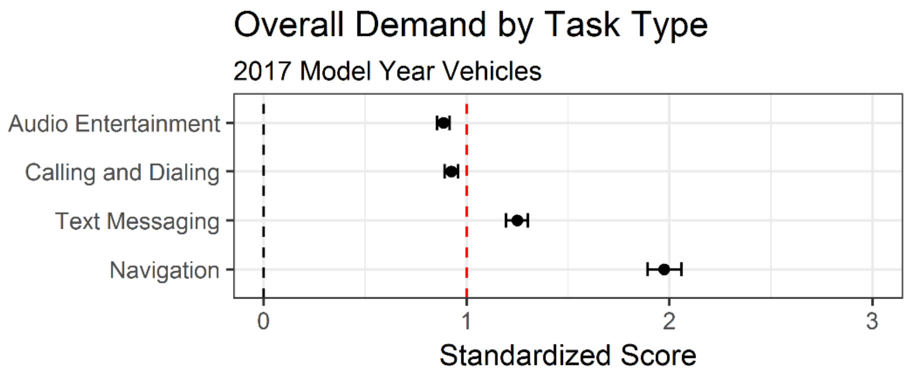

Figure 7. Overall demand as a function of task type for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the high demand referent tasks. Error bars represent 95% confidence intervals.

Effects of Mode of Interaction

Figures 8-12 present the workload associated with the *three modes of interaction* evaluated in the on-road testing. Figure 8 presents the cognitive demand, Figure 9 presents the visual demand, Figure 10 presents the subjective demand, and Figure 11 presents the task interaction time. The overall demand is presented in Figure 12.

Cognitive demand was derived using Equation 1. A comparison of linear mixed effects models with and without Modality indicated that Modality was a significant predictor of cognitive demand ($\chi^2(2) = 60.2$, $p < .001$). Inspection of Figure 8 shows that the cognitive demand of each modality of interaction was greater than the N-back task. The relative ordering placed the auditory vocal interactions as less cognitively demanding than the center stack interactions, which were less demanding than the center console interactions.

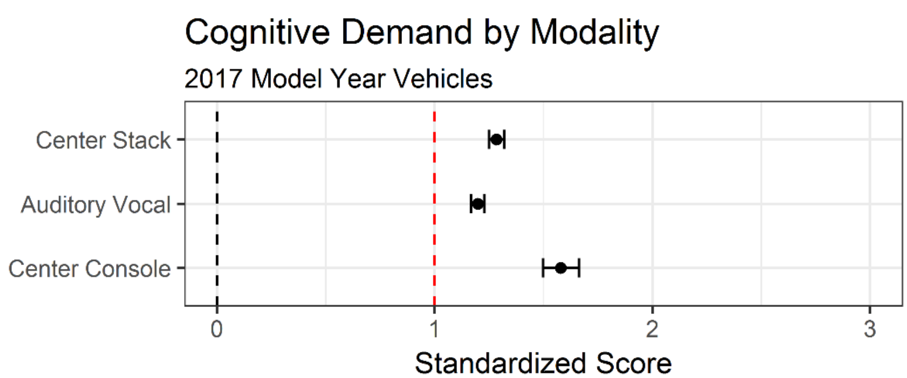

Figure 8. Cognitive demand as a function of mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the N-bask referent task. Error bars represent 95% confidence intervals.

Visual demand was derived using Equation 2. A comparison of linear mixed effects models with and without Modality indicated that Modality was a significant predictor of visual demand ($\chi^2(2) = 1059$, $p < .001$). Inspection of Figure 9 shows that the visual demand was significantly lower than the SuRT task for the auditory vocal interactions, as expected, but was higher significantly higher than the SuRT task for the center console and center stack interactions. Center console interactions were less visually demanding than center stack interactions.

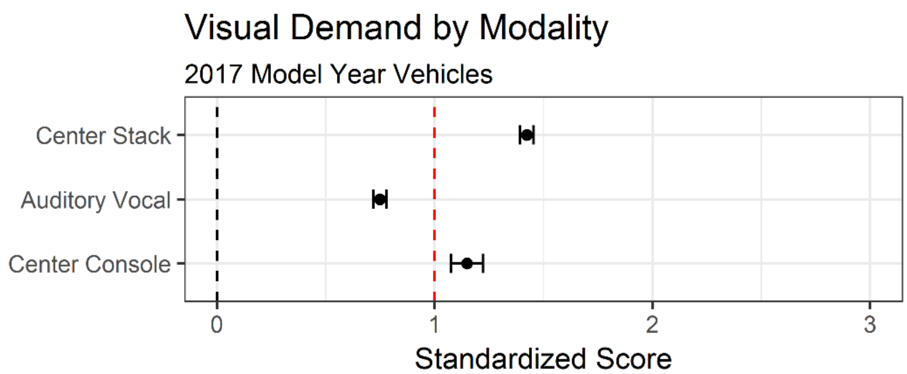

Figure 9. Visual demand as a function of mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the SuRT task. Error bars represent 95% confidence intervals.

Subjective demand was derived using Equation 3. A comparison of linear mixed effects models with and without Modality indicated that Modality was a significant predictor of subjective demand ($\chi^2(2) = 459$, $p < .001$). Inspection of Figure 10 shows that the subjective demand was lower than the high demand referent tasks. Auditory vocal interactions were subjectively less demanding than center console and center stack interactions. Center console interactions were subjectively less demanding than center stack interactions.

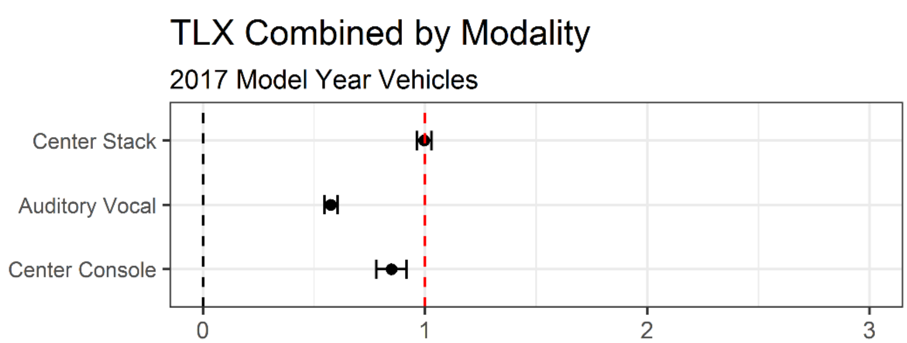

Figure 10. Subjective demand as a function of mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the average demand of the N-back and SuRT tasks. Error bars represent 95% confidence intervals.

The interaction time was derived using Equation 4. A comparison of linear mixed effects models with and without Modality indicated that Modality was a significant predictor of interaction time ($\chi^2(2) = 1079$, $p < .001$). Inspection of Figure 11 shows that center stack interactions took significantly less time than the 24-second standard and auditory vocal tasks took significantly more time than the 24-second standard. Center console interaction time did not differ significantly from the 24-second standard.

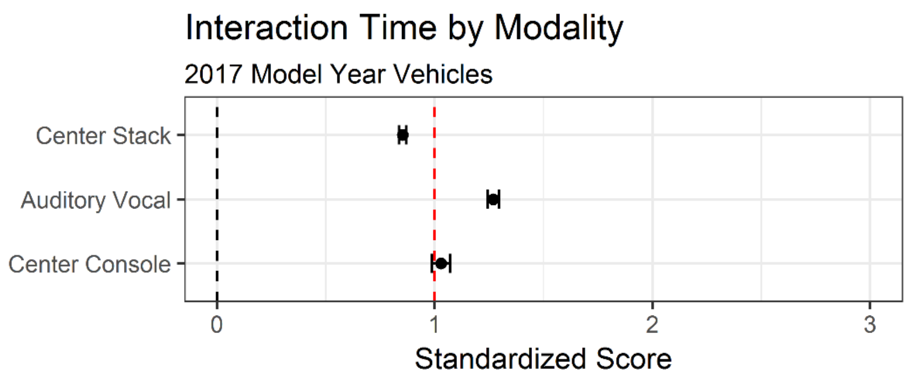

Figure 11. Interaction time as a function of mode of interaction for the on-road assessment. The dashed vertical red line represents the 24-second task interaction standard. Error bars represent 95% confidence intervals.

Finally, overall demand, derived using Equation 5 and presented in Figure 12, shows that all the tasks exceeded the standardized high workload referent represented by the red vertical line. A comparison of linear mixed effects models with and without Modality indicated that Modality was a significant predictor of overall demand ($\chi^2(2) = 11.08$, $p < .01$). Of the three modes of interaction evaluated, center stack interactions were the easiest to perform. Auditory vocal interactions were more demanding than center stack interactions. The center console was the most demanding mode of interaction that we evaluated.

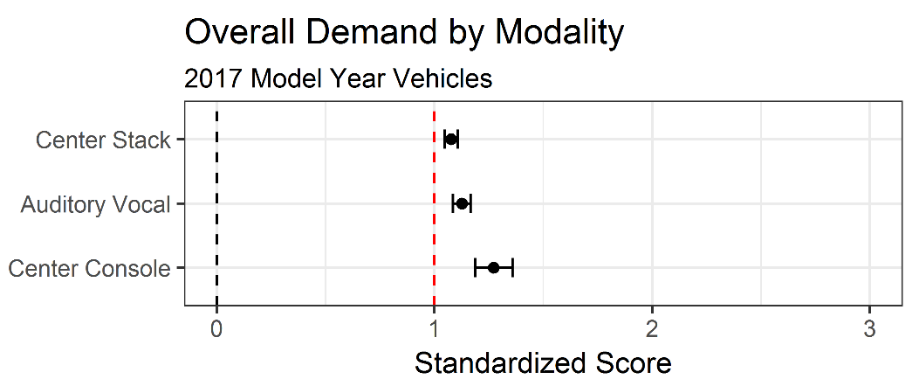

Figure 12. Overall demand as a function of mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the high demand referent. Error bars represent 95% confidence intervals.

Effects of Task Type and Mode of Interaction

Figures 13-17 present the workload broken down by task type and mode of interaction. Figure 13 presents the cognitive demand, Figure 14 presents the visual demand, Figure 15 presents the subjective demand, and Figure 16 presents the interaction time by task type. The overall demand is presented in Figure 17.

Cognitive demand was derived using Equation 1, and inspection of Figure 13 shows the cognitive demand broken down by task type and modality of interaction. A linear mixed effects models analysis indicated that the interaction between task type and modality was not significant ($\chi^2(6) = 9.1$, $p = .166$).

In all cases, the cognitive demand was significantly greater than the N-back task. Cognitive demand was lowest for the auditory vocal interactions followed by center stack interactions followed by center console interactions.

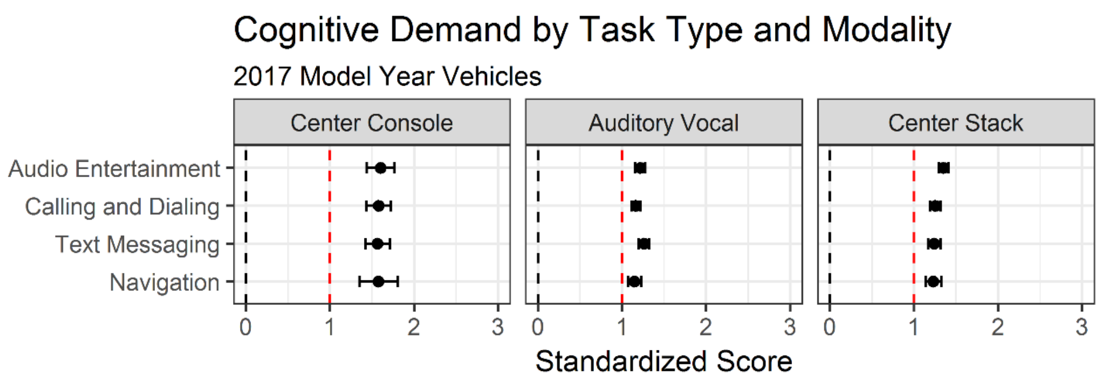

Figure 13. Cognitive demand as a function of task type and mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the N-bask referent task. Error bars represent 95% confidence intervals.

Visual demand was derived using Equation 2 and inspection of Figure 14 shows that the visual demand tended to be lower for task types supported by auditory vocal interactions. The other task type by modes of interaction tended to have visual demand greater than the SuRT task. A linear mixed effects models analysis indicated that the interaction between task type and modality was predictive of visual demand ($\chi^2(6) = 143$, $p < .001$).

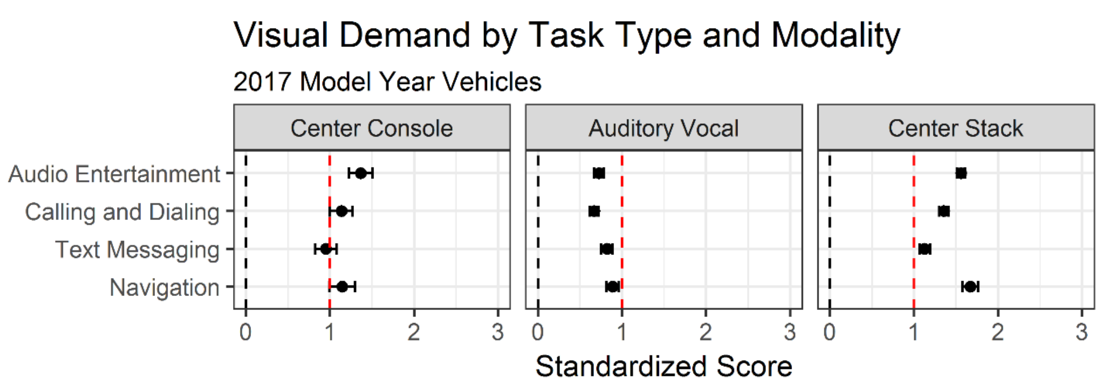

Figure 14. Visual demand as a function of task type and mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the SuRT task. Error bars represent 95% confidence intervals.

Subjective demand was derived using Equation 3 and inspection of Figure 15 shows that the subjective demand of the auditory vocal and center console modes of interaction was lower than the high demand referent tasks and the center stack interactions did not differ from the high demand referent. The relative ordering of task types placed audio entertainment as less demanding than center console interactions, which were less demanding than center stack interactions. A linear mixed effects models analysis indicated that the interaction between task type and modality was predictive of subjective demand ($\chi^2(6) = 143$, $p < .001$).

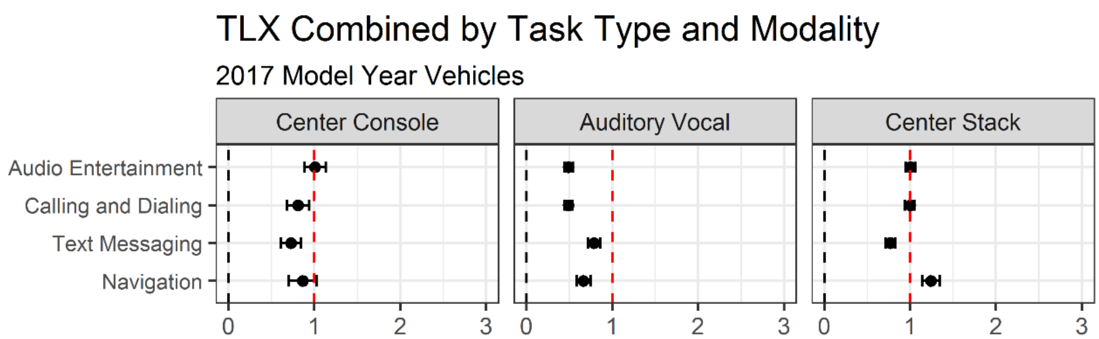

Figure 15. Subjective demand as a function of task type and mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the average demand of the N-back and SuRT tasks. Error bars represent 95% confidence intervals.

The interaction time was derived using Equation 4 and is presented in Figure 16. The center stack interactions generally took less time than the center console interactions, which took less time to perform than the auditory vocal interactions. In all cases, the navigation task took the longest time to complete and this was most pronounced with the auditory vocal interaction modality. A linear mixed effects models analysis indicated that the interaction between task type and modality was predictive of interaction time ($\chi^2(6) = 1051$, $p < .001$).

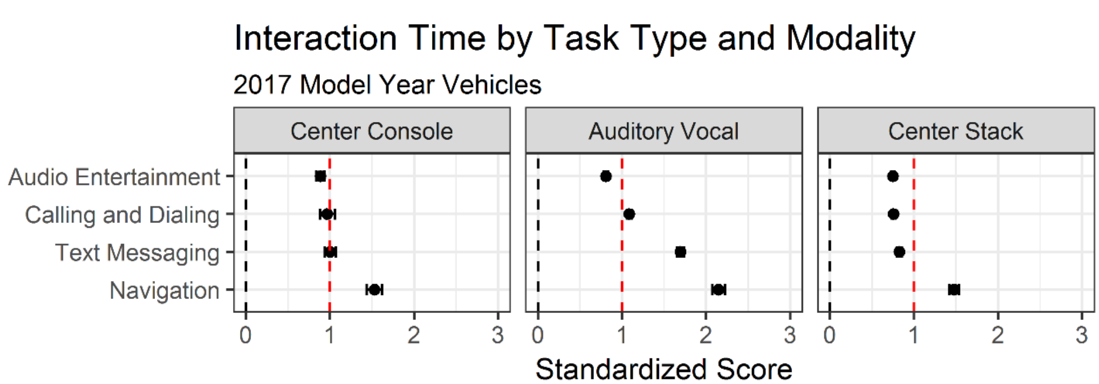

Figure 16. Interaction time as a function of task type and mode of interaction for the on-road assessment. The dashed vertical red line represents the 24-second task interaction standard. Error bars represent 95% confidence intervals.

Finally, overall demand, derived using Equation 5, is presented in Figure 17. A linear mixed effects models analysis indicated that the interaction between task type and modality was predictive of overall demand ($\chi^2(6) = 445$, $p < .001$). Of the four tasks by three modes of interaction evaluated, the navigation tasks were consistently highest in demand, and there does not appear to be a configuration that is not excessive. The task types performed using the center stack were less variable and had lower overall demand than similar task types performed using the center console, indicating that the former interface was superior to the latter. The audio vocal interactions exhibited the greatest range of overall demand by task type, with the lowest demand for the audio entertainment and calling and dialing task types and the highest demand for the text messaging and navigation task types. The best mode of interaction for audio entertainment and calling and dialing was the auditory vocal interface, whereas the best mode of interaction for text messaging was the center stack interface. The greater overall demand of the auditory vocal text messaging interactions was driven by the longer task interaction times. There was no good mode of interaction for the navigation task.

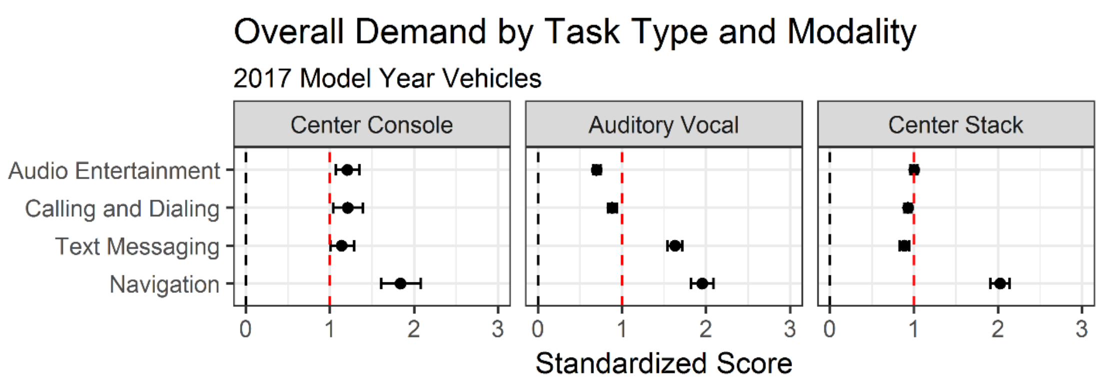

Figure 17. Overall demand as a function of task type and mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the high demand referent task. Error bars represent 95% confidence intervals.

Effects of Vehicle

Figures 18-22 present the workload associated with the vehicles evaluated in the on-road testing. Figure 18 presents the cognitive demand, Figure 19 presents the visual demand, Figure 20 presents the subjective demand, and Figure 21 presents the task interaction time. The integrated demand is presented in Figure 22.

Cognitive demand was derived using Equation 1. A comparison of linear mixed effects models with and without Vehicle indicated that Vehicle was a significant predictor of cognitive demand ($\chi^2(29) = 184$, $p < .001$). Inspection of Figure 18 shows that the cognitive demand varied as a function of vehicle. The Chevrolet Equinox LT, Chrysler 300C, Ford Fusion Titanium, Ford Mustang GT Premium Convertible, Nissan Armada SV, and Tesla Model S 75 had a cognitive demand score that did not significantly differ from the N-back task. The remaining vehicles all had cognitive demand that was greater than the N-back task.

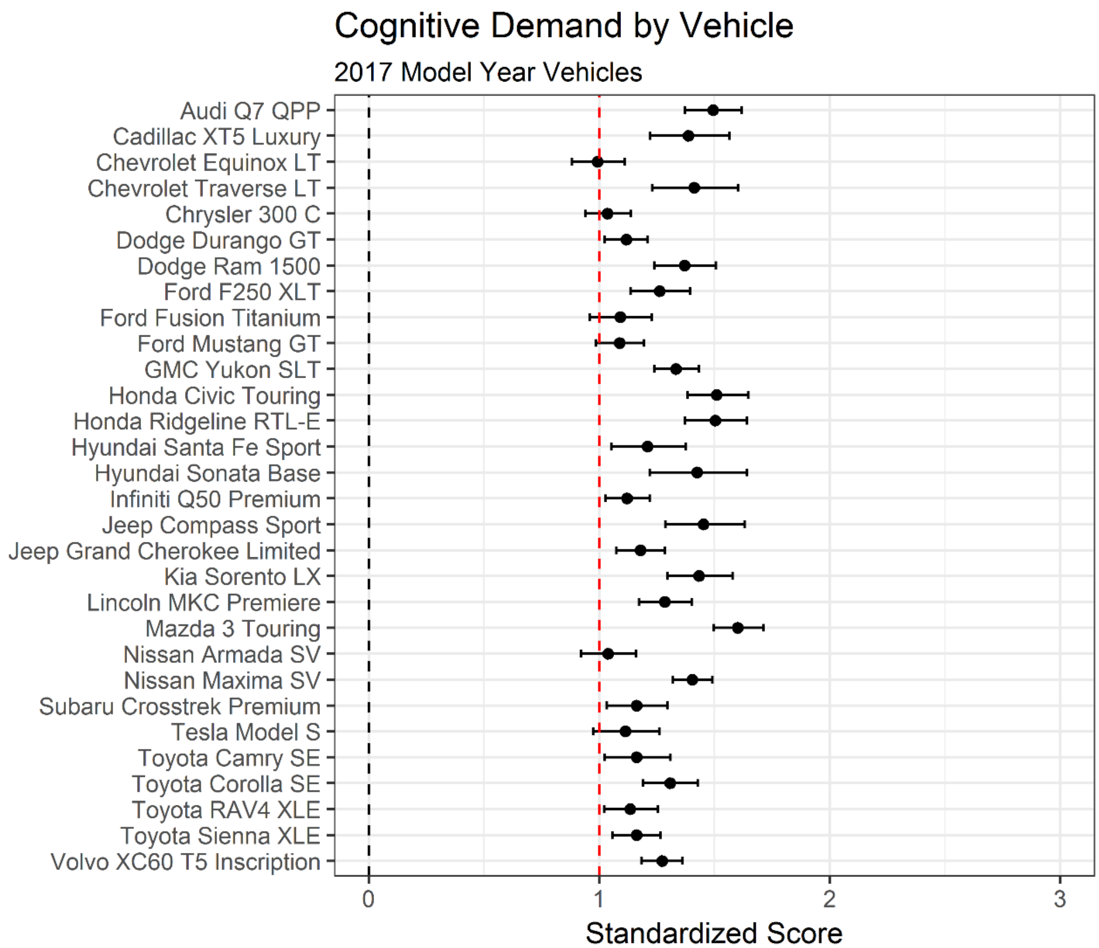

Figure 18. Cognitive demand as a function of vehicle for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the N-bask referent task. Error bars represent 95% confidence intervals.

Visual demand was derived using Equation 2. A comparison of linear mixed effects models with and without Vehicle indicated that Vehicle was a significant predictor of visual demand ($\chi^2(29) = 346$, p < .001). Inspection of Figure 19 shows that the visual demand was significantly lower than the SuRT task for the Chevrolet Equinox LT, Ford F-250 XLT, Jeep Compass Sport, Jeep Grand Cherokee, Mazda3 Touring, and Toyota Sienna XLE. Visual demand was significantly greater than the SuRT task for the Cadillac XT5 Luxury, Chrysler 300C, Dodge Durango GT, Ford Mustang GT Premium Convertible, GMC Yukon SLT, Honda Civic Touring, Honda Ridgeline RTL-E, Hyundai Santa Fe Sport, Lincoln MKC Premiere, Nissan Armada SV, Tesla Model S 75, and the Volvo XC60 T5 Inscription.

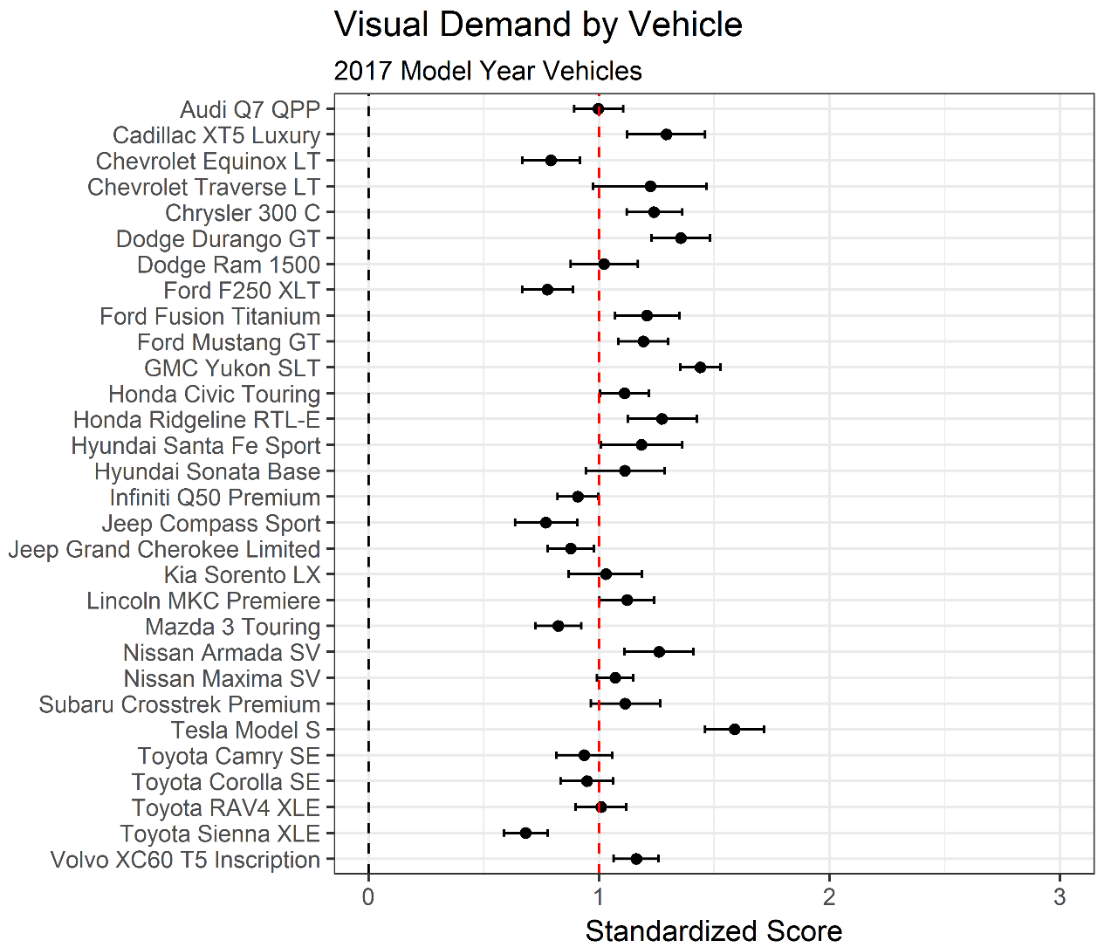

Figure 19. Visual demand as a function of vehicle for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the SuRT task. Error bars represent 95% confidence intervals.

Subjective demand was derived using Equation 3. A comparison of linear mixed effects models with and without Vehicle indicated that Vehicle was a significant predictor of subjective demand ($\chi^2(29) = 167$, $p < .001$). Inspection of Figure 20 shows that the subjective demand varied as a function of vehicle. The Chevrolet Traverse LT, Ford Mustang GT Premium Convertible, Honda Ridgeline RTL-E, Kia Sorento LX, Subaru Crosstrek Premium, Tesla Model S 75, and Volvo XC60 T5 Inscription did not differ from the average of the high demand referent tasks. All other vehicles scored significantly below the high demand referent.

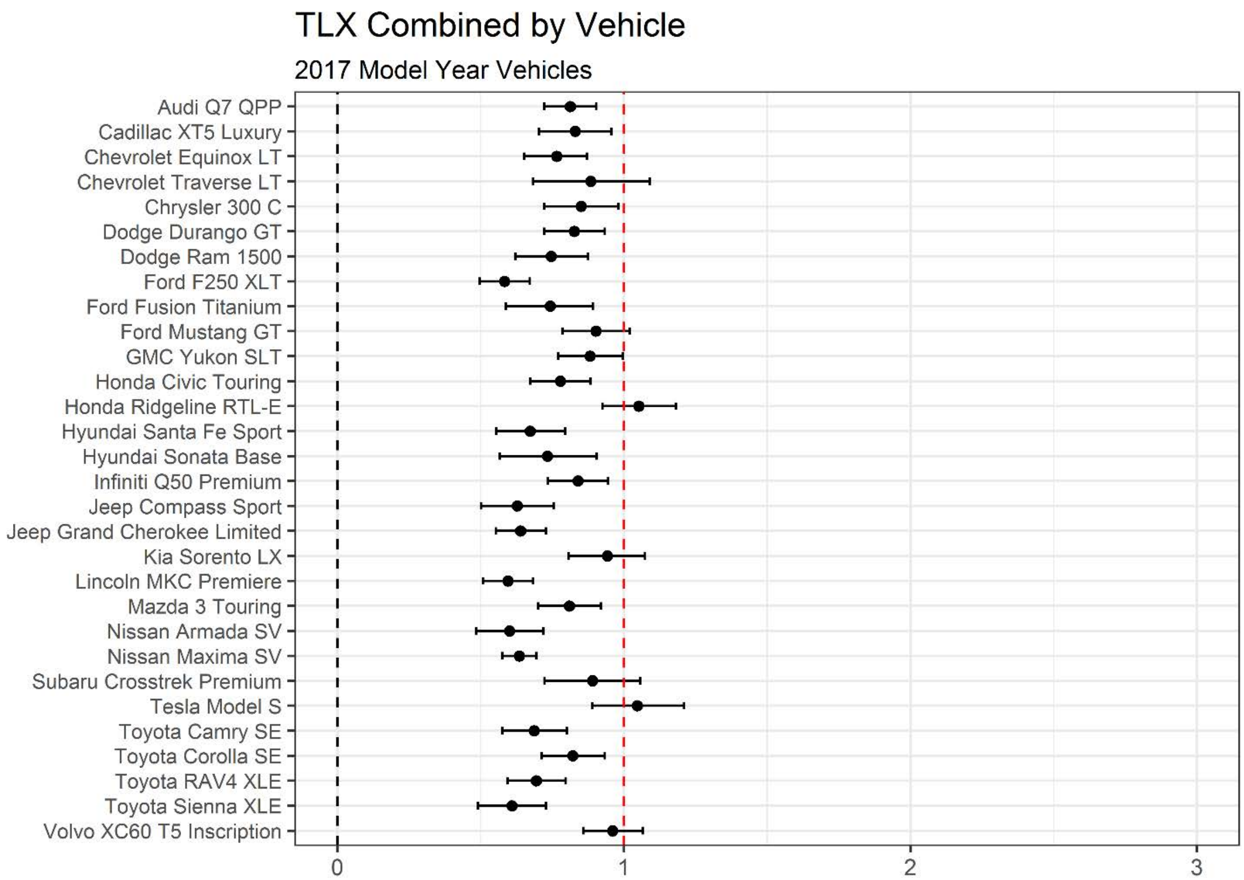

Figure 20. Subjective demand as a function of vehicle for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the average demand of the N-back and SuRT tasks. Error bars represent 95% confidence intervals.

Task interaction time was derived using Equation 4. A comparison of linear mixed effects models with and without Vehicle indicated that Vehicle was a significant predictor of interaction time ($\chi^2(29) = 744$, $p < .001$). Inspection of Figure 17 showed that this measure varied as a function of vehicle. Vehicles that significantly exceeded the 24-second standard were the Audi Q7 3.0T Premium Plus, Chrysler 300C, Dodge Durango GT, Ford Mustang GT Premium Convertible, Mazda3 Touring, Nissan Armada SV, Subaru Crosstrek Premium, Honda Civic Touring, Honda Ridgeline RTL-E and the Volvo XC60 T5 Inscription.

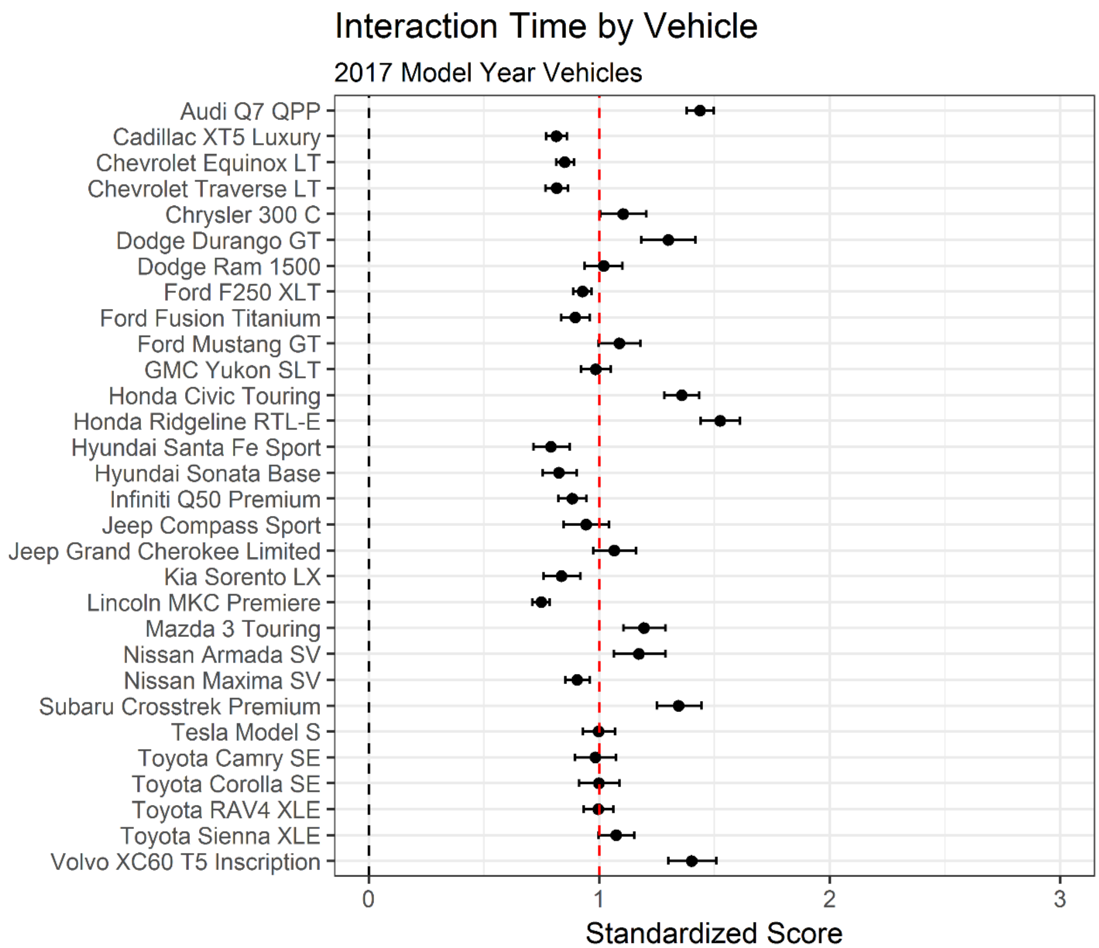

Figure 21. Interaction time as a function of vehicle for the on-road assessment. The dashed vertical red line represents the 24-second task interaction referent. Error bars represent 95% confidence intervals.

Finally, overall demand, derived using Equation 5 and presented in Figure 22, shows that the majority of vehicles were at or exceeded the standardized high workload benchmark represented by the red vertical line. A comparison of linear mixed effects models with and without Vehicle indicated that Vehicle was a significant predictor of overall demand ($\chi^2(29) = 455$, $p < .001$). In the figure, vehicles are ordered by increasing levels of overall demand and there is a noticeable positive skew in the ratings. Seven of the vehicles received an overall rating significantly below 1.0 (i.e., a moderate level of overall demand); 11 vehicles received a score that did not differ from the high demand referent (i.e., a high overall demand score), and 12 vehicles scored significantly above the high demand referent (i.e., a very high overall demand score).

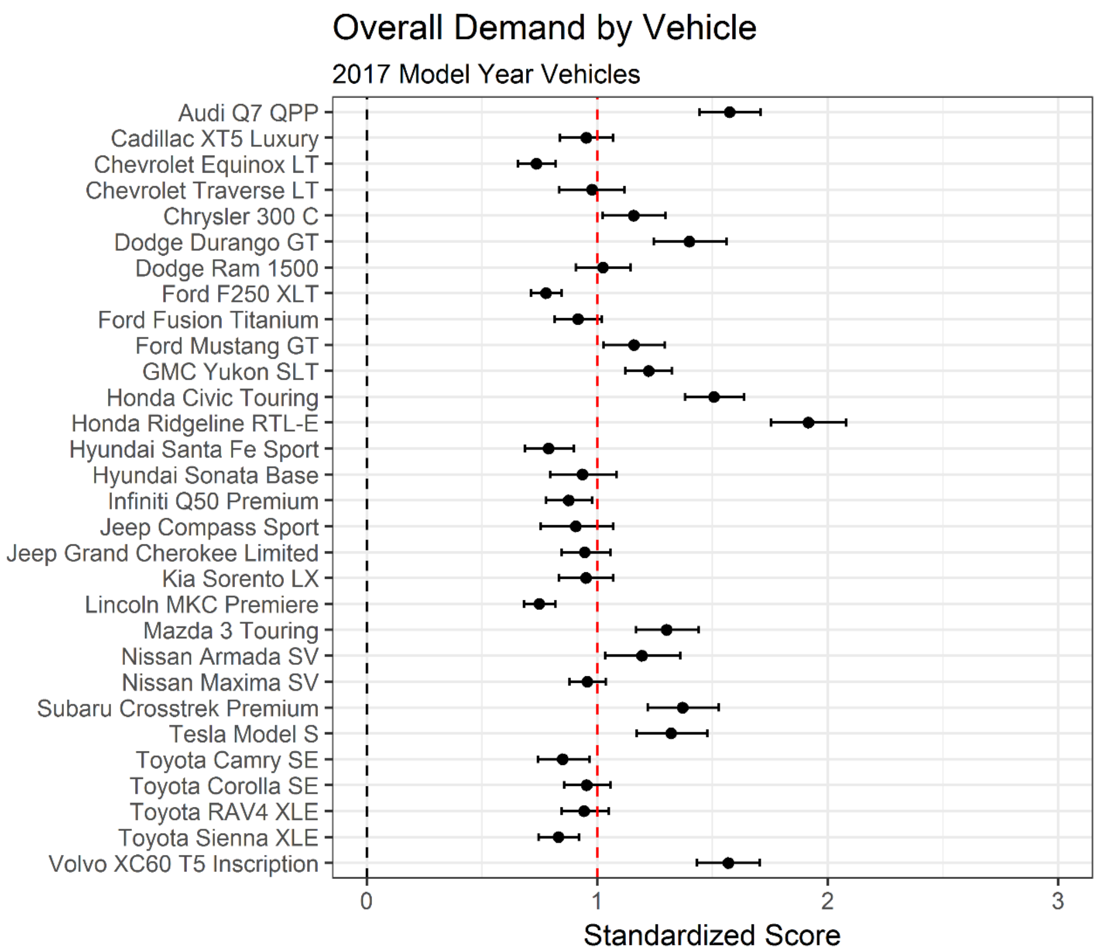

Figure 22. Overall demand as a function of vehicle for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the high demand referent. Error bars represent 95% confidence intervals.

Effects of Vehicle and Task Type

Figures 23-27 present the demand broken down by vehicle and task. Figure 23 presents the cognitive demand, Figure 24 presents the visual demand, Figure 25 presents the subjective demand, and Figure 26 presents the interaction time by task. The overall demand is presented in Figure 27. Not all tasks were available in all vehicles we tested. Calling and dialing and audio entertainment were available in all vehicles. The ability to use the IVIS to interact with text messaging was available in most vehicles (except the Cadillac XT5 Luxury, Chevrolet Traverse LT, Hyundai Santa Fe Sport, Hyundai Sonata Base, Kia Sorento LX, Nissan Armada SV, Subaru Crosstrek Premium, and the Tesla Model S 75). The 12 vehicles that supported navigation were the Audi Q7 3.0T Premium Plus, Chrysler 300C, Dodge Durango GT, Ford Mustang GT Premium Convertible, GMC Yukon SLT, Honda Civic Touring, Honda Ridgeline RTL-E, Jeep Grand Cherokee Limited, Nissan Armada SV, Nissan Maxima SV, Tesla Model S 75, and the Volvo XC60 T5 Inscription.

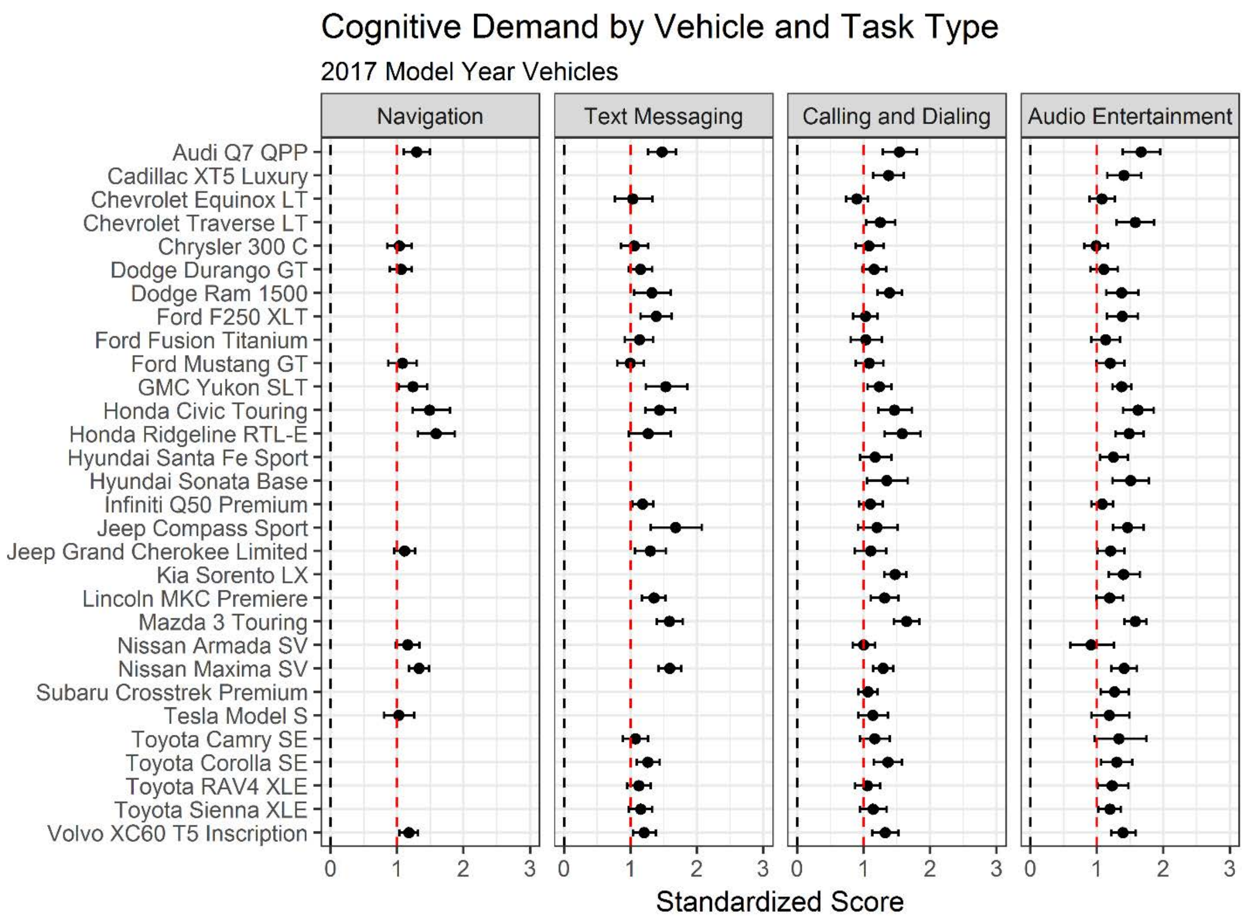

Figure 23. Cognitive demand as a function of vehicle and task for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the N-bask referent task. Error bars represent 95% confidence intervals.

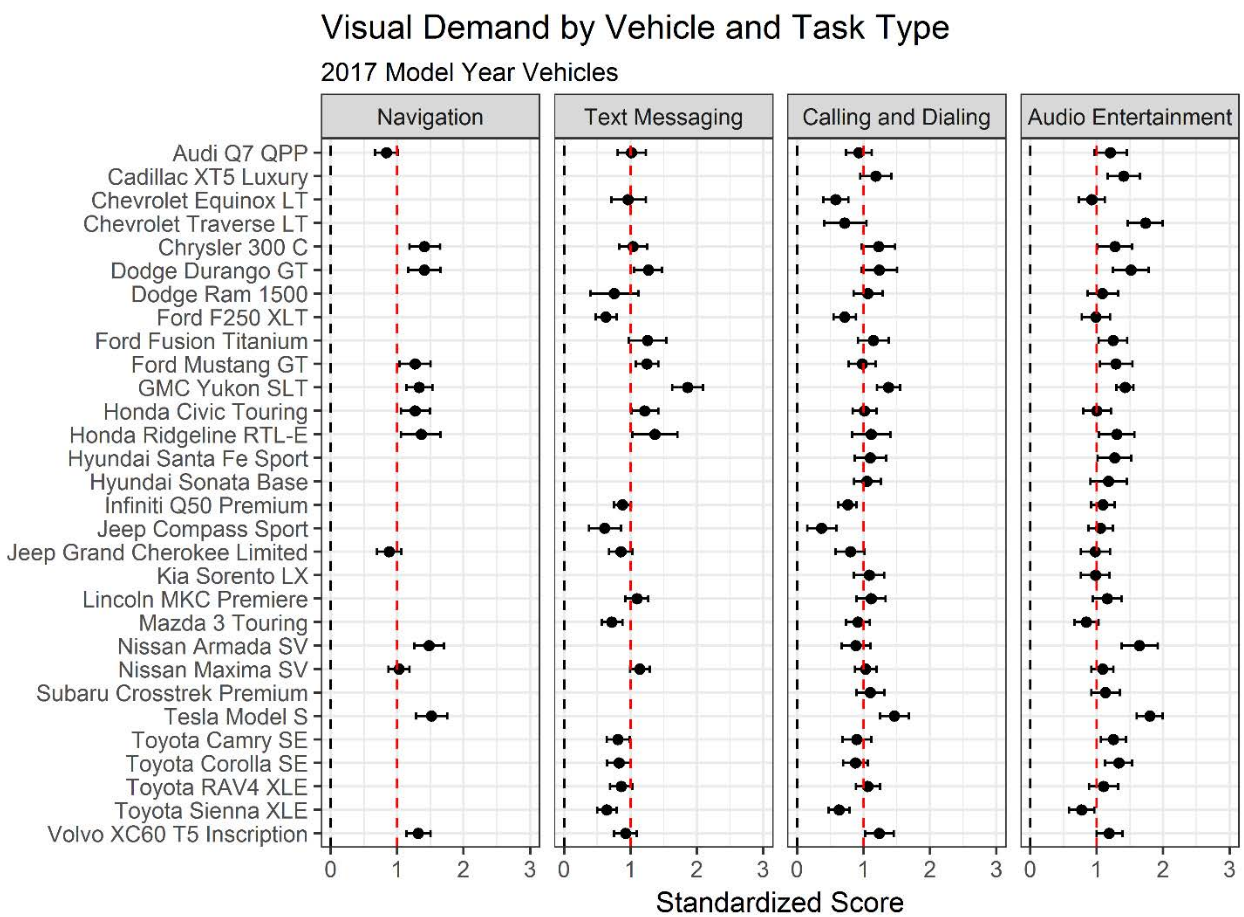

Figure 24. Visual demand as a function of vehicle and task type for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the SuRT task. Error bars represent 95% confidence intervals.

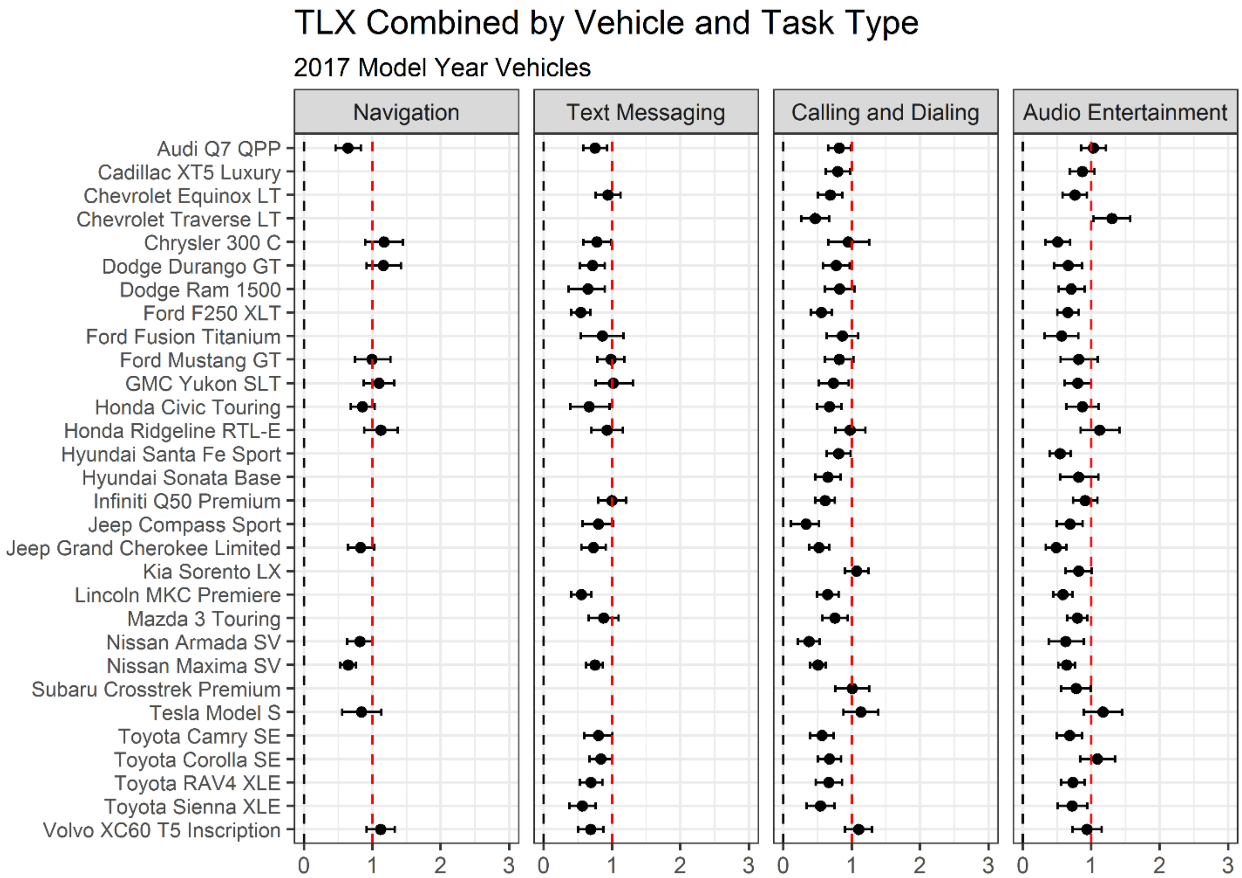

Figure 25. Subjective demand as a function of vehicle and task type for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the average demand of the N-back and SuRT tasks. Error bars represent 95% confidence intervals.

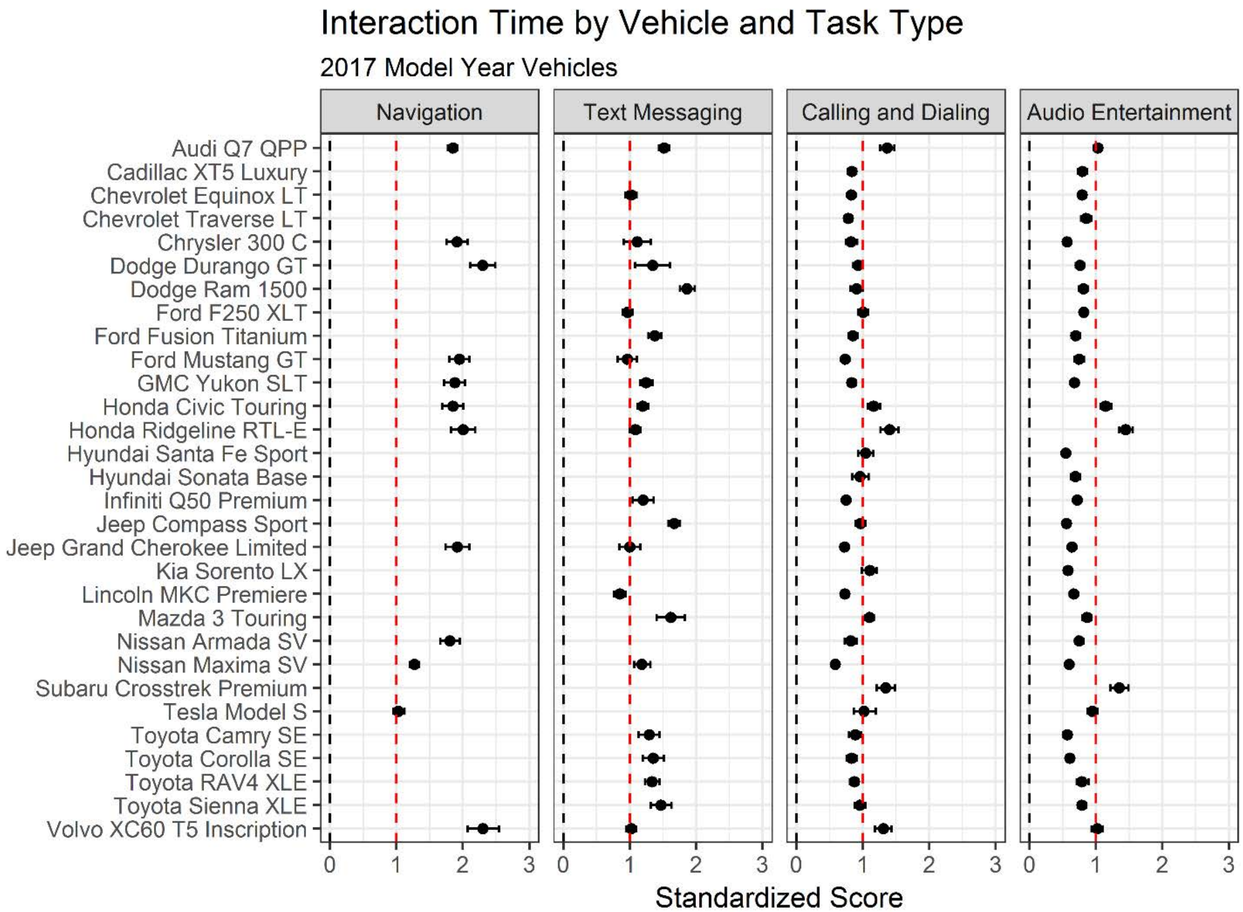

Figure 26. Interaction time as a function of vehicle and task type for the on-road assessment. The dashed vertical red line represents the 24-second task interaction referent. Error bars represent 95% confidence intervals.

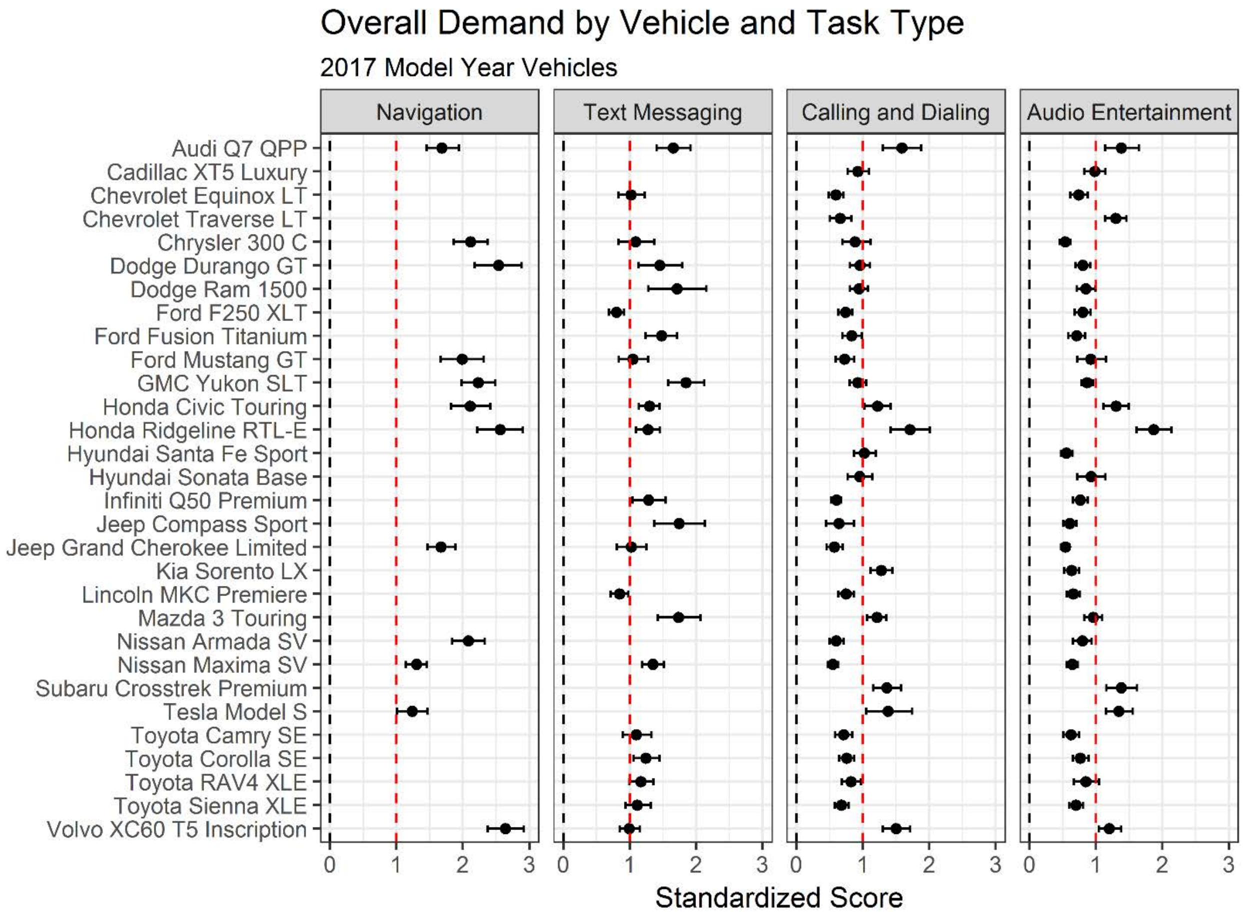

Figure 27. Overall demand as a function of vehicle and task type for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the high demand referent. Error bars represent 95% confidence intervals.

Effects of Vehicle and Modality of Interaction

Figures 28-32 present the demand broken down by vehicle and mode of interaction. Figure 28 presents the cognitive demand, Figure 29 presents the visual demand, Figure 30 presents the subjective demand, and Figure 31 presents the interaction time by task. The overall demand is presented in Figure 32. Not all modes of interaction were available in all vehicles we tested. Audio vocal interactions were available in all vehicles we tested. Center stack interactions were available in all vehicles except the Audi Q7 3.0T Premium Plus, and the Mazda3 Touring. Center console interactions were available only in the Audi Q7 3.0T Premium Plus, Infinity Q50 Premium, Mazda3 Touring, and the Nissan Maxima SV vehicles.

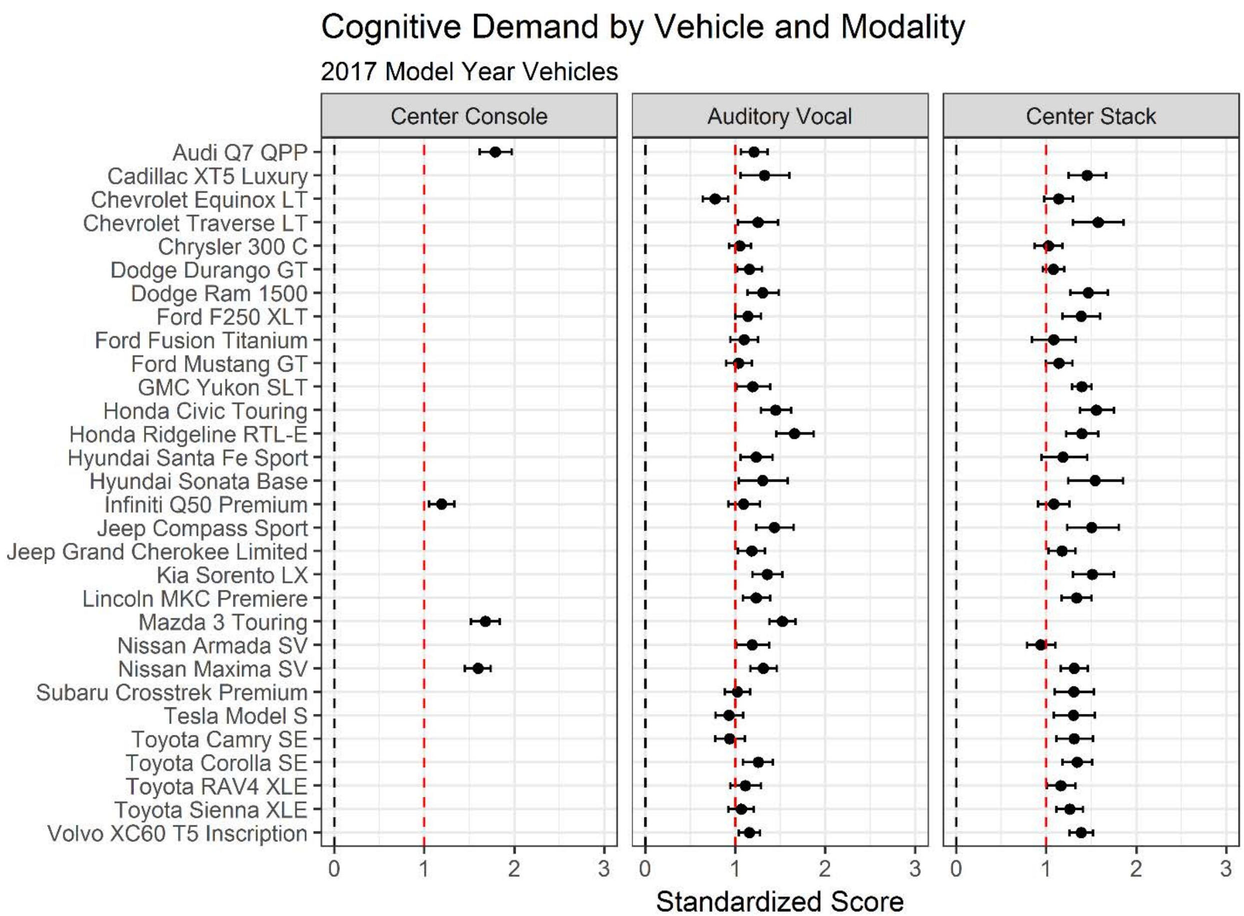

Figure 28. Cognitive demand as a function of vehicle and mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the N-bask referent task. Error bars represent 95% confidence intervals.

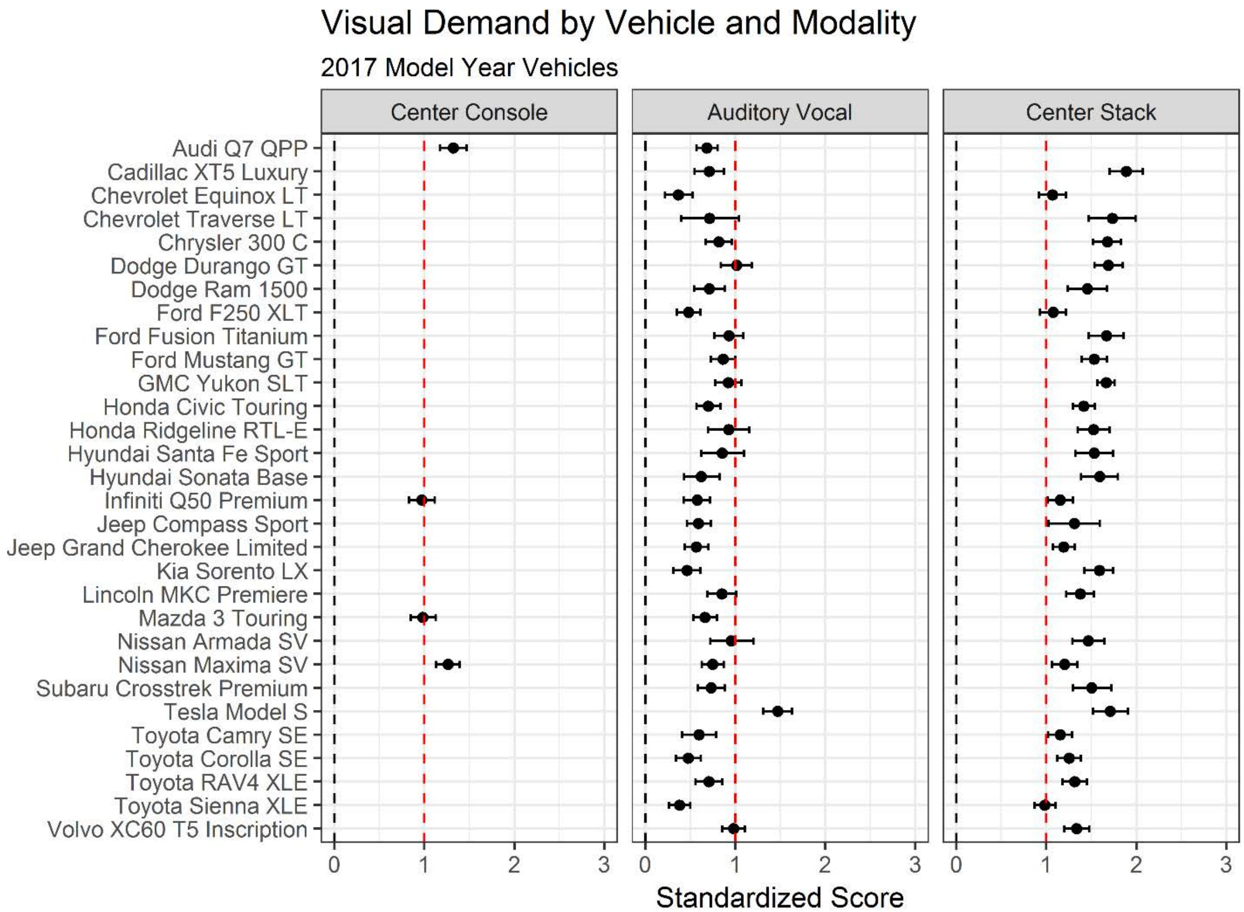

Figure 29. Visual demand as a function of vehicle and mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the performance on the SuRT task.

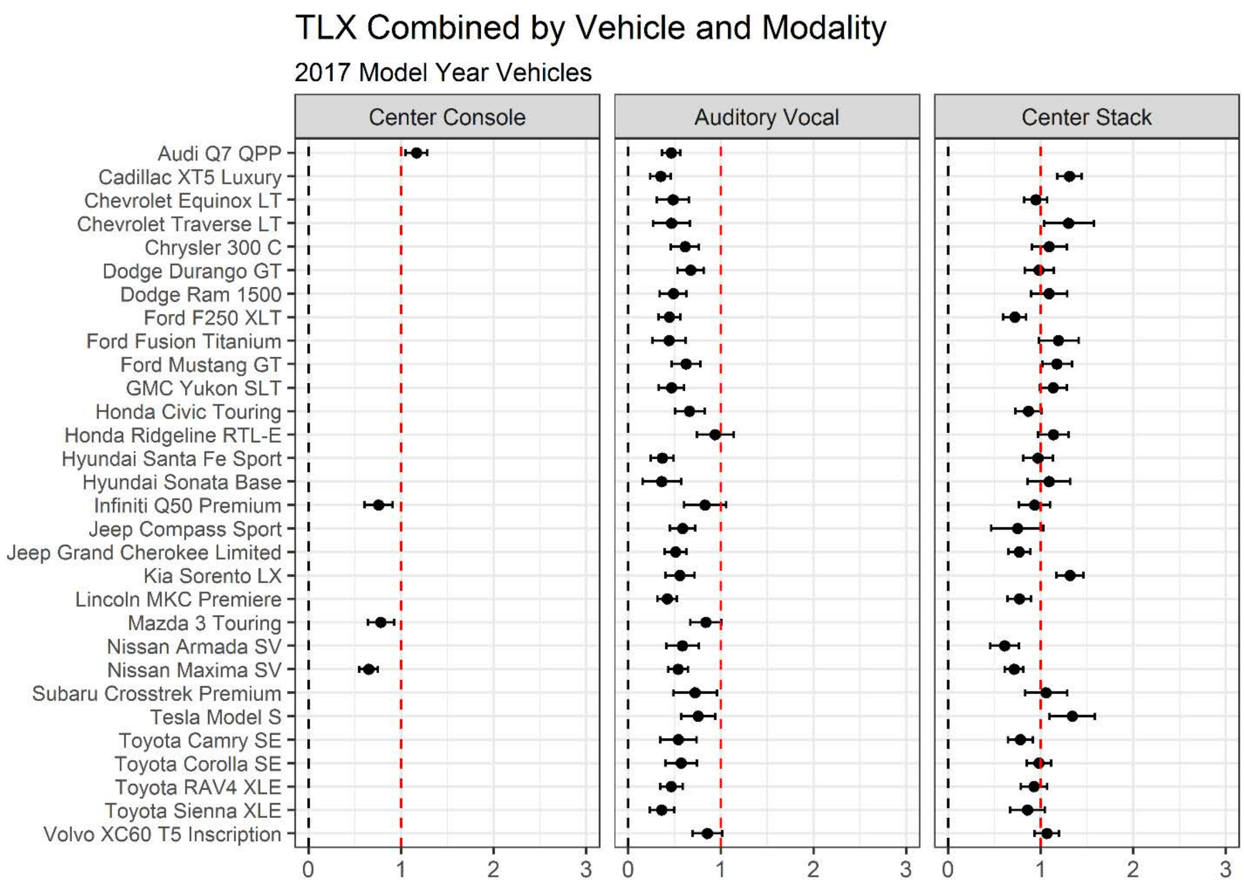

Figure 30. Subjective demand as a function of vehicle and mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the average demand of the N-back and SuRT tasks. Error bars represent 95% confidence intervals.

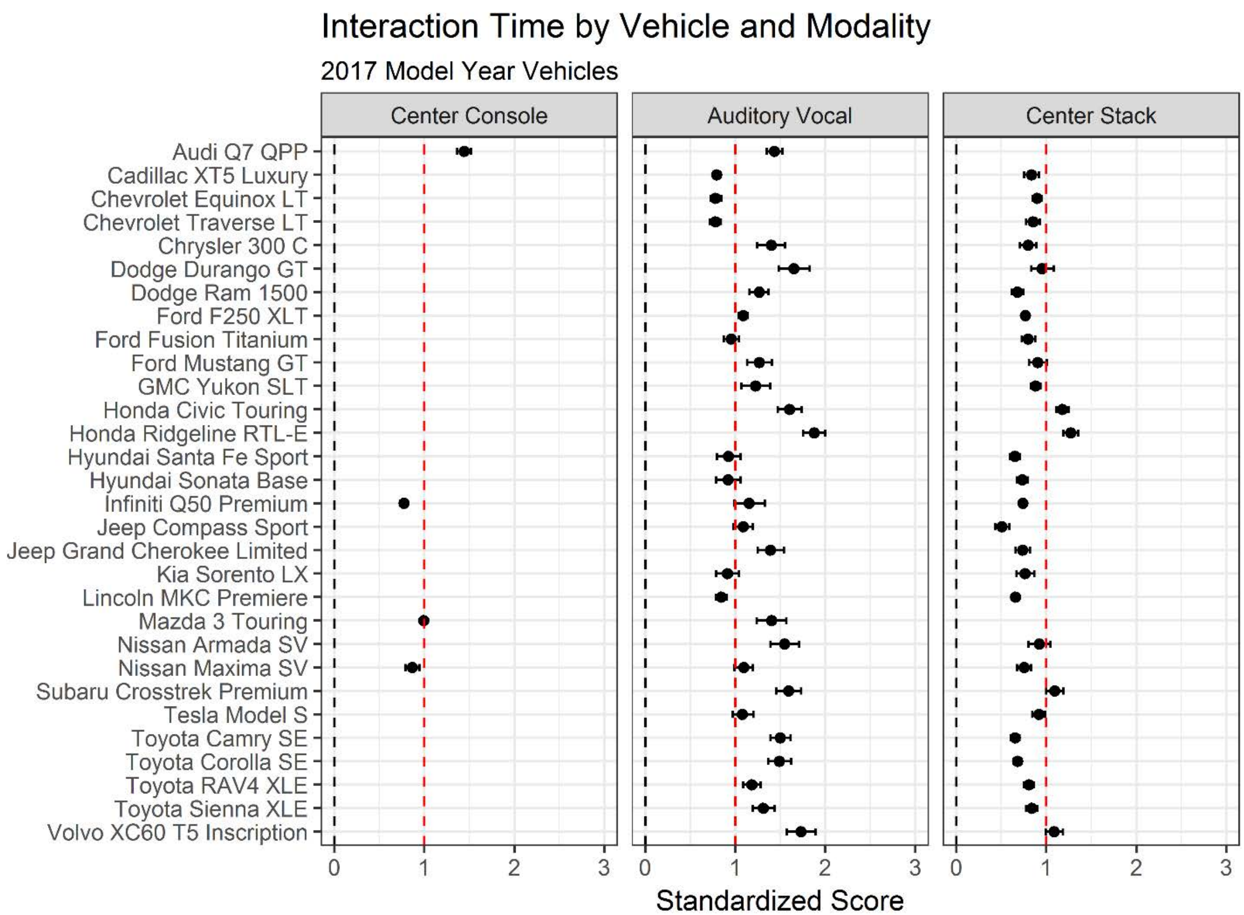

Figure 31. Interaction time as a function of vehicle and mode of interaction for the on-road assessment. The dashed vertical red line represents the 24-second task interaction referent. Error bars represent 95% confidence intervals.

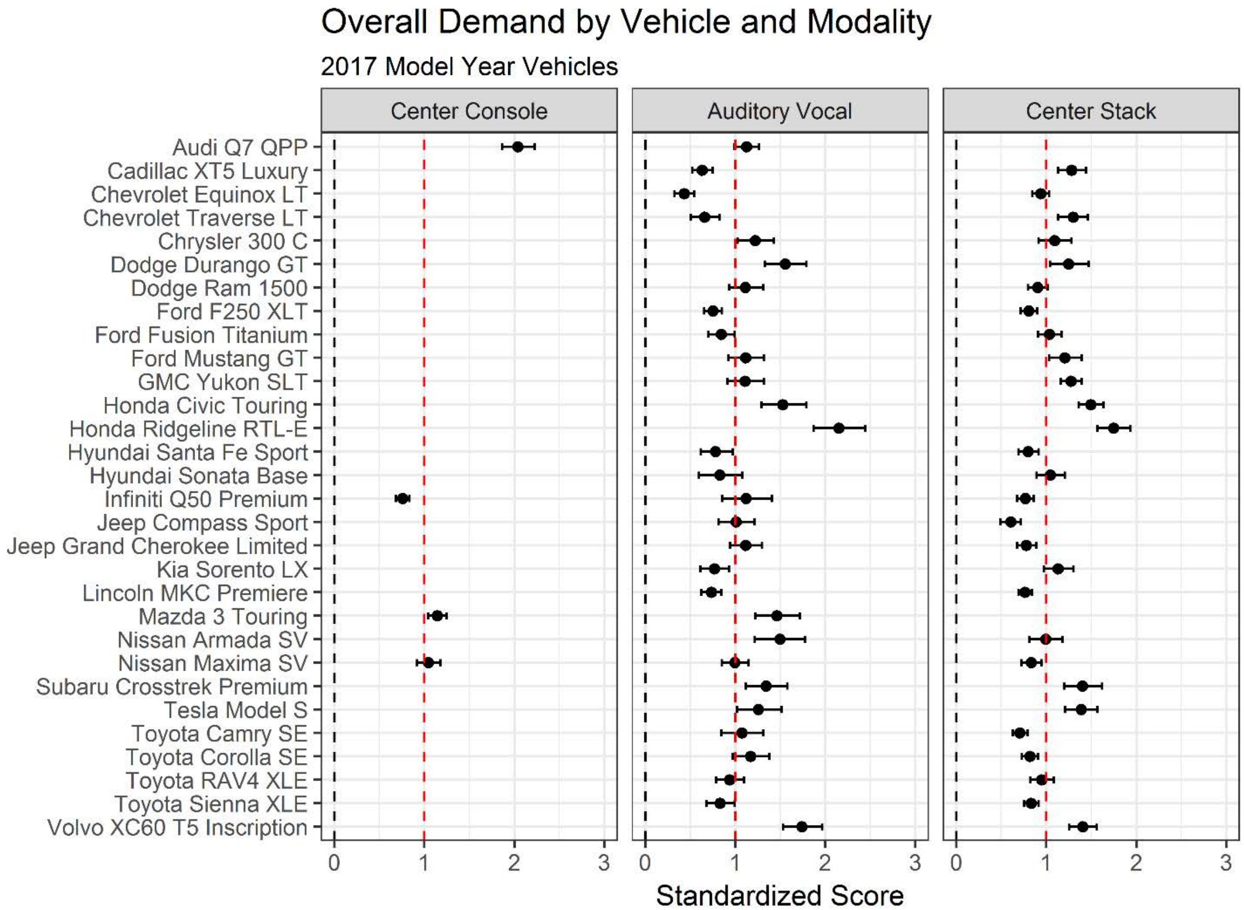

Figure 32. Overall demand as a function of vehicle and mode of interaction for the on-road assessment. The dashed vertical black line represents single-task performance and the dashed vertical red line represents the high demand referent. Error bars represent 95% confidence intervals.

DISCUSSION

2017 model-year automobiles provide an unprecedented number of features and functions that allow motorists to perform a variety of secondary tasks unrelated to the primary task of driving. Surprisingly little is known about how these complex multimodal IVIS interactions impact a driver's workload. Given the ubiquity of these systems, the current research used cutting-edge methods to address three interrelated questions concerning this knowledge gap.

First, are some task types more impairing than others? The answer to this question can be seen most directly in Figure 7, which plots overall demand as a function of the IVIS task types we tested on the road. In that figure, the overall workload associated with audio entertainment and calling and dialing task types were easier than the high demand referent, standardized as a score of 1.0 and indicated by a red vertical line. Text messaging and the navigation task types were harder than the high demand referent. The IVIS task types differed in terms of demand, with audio entertainment task type being statistically equivalent to the calling and dialing task type (the two most universal of IVIS tasks available in all 2017 model-year automobiles we tested). Text messaging, an IVIS feature found in 22 out of 30 vehicles we tested, was associated with a significantly higher level of demand than the former task types and was significantly higher than the high demand referent. Most demanding of all was destination entry for navigation, an IVIS feature that was available in 12 out of 30 of the vehicles we evaluated. The navigation task type had an overall demand that was more than two times that of the high demand referent.

One critical factor for the high workload ratings was the interaction time. The shortest interaction times were associated with audio entertainment. Calling and dialing took significantly longer than the selection of music. Texting took an average of 30 seconds and destination entry for navigation took an average of 40 seconds. Clearly, the latter two task types divert the driver's attention from the road for far too long. For example, at 25 mph drivers would travel just under 1,500 feet while using the IVIS for entering destinations, and several of the navigation systems that were tested took considerably longer than the 40-second average.

Of note were the subjective ratings, which tracked reasonably well with the measures of cognitive and visual demand, but not with interaction time. For example, the subjective rating of demand for the navigation task did not differ from the audio entertainment task, despite a more than 2:1 difference in interaction time. These data call into question assumptions that motorists may self-regulate their secondary-task behavior (see Sanbonmatsu et al., 2016). That is, from the driver's *subjective* perspective, the two tasks were very similar, whereas the measures of overall demand associated with *objective* measures tells a very different story.

Second, are some modes of interaction more distracting than others? The answer to this question can be seen in Figure 12, which plots overall demand as a function of the mode of IVIS interaction. The overall workload associated with each mode of IVIS interaction was greater than the high workload referent, standardized as a score of 1.0 and indicated by a red vertical line. Interactions using the center stack were significantly less demanding than auditory vocal interactions, which were less demanding than center console interactions. Interestingly, using voice-based commands to control IVIS functions resulted in significantly lower levels of visual demand than the SuRT task. By design, auditory-vocal interfaces allow the driver to keep their eyes on the road while interacting with the IVIS, but with this type of interaction motorists are significantly less likely to see what they are looking at (Strayer, Drews, & Johnston, 2003). However, the benefits of reduced visual demand are offset by longer interaction times. Auditory vocal interactions took significantly longer than any other IVIS interaction (an average of 30 seconds in our testing).

Some combinations of the task type by mode of IVIS interaction were less demanding than others. Inspection of Figure 17 shows that using voice commands to select music or place phone calls was associated with lower levels of workload than for the other interaction modalities. By contrast, using the touch screen on the center stack was the least demanding way to send a text message (and conversely, using voice commands was the most demanding interaction modality for texting). Although entering destinations while the vehicle was in motion was associated with very high levels of demand, these interactions were somewhat less demanding when performed using the touch screen on the center stack. Center console interactions were supported in four out of the 30 vehicles that we tested. This included the use of a rotary dial and/or writing pad as a mode of input. The center console was significantly worse that the other interaction modalities, except for the navigation task where all modes of interaction fared poorly. This may stem from imperfect stimulus-response mapping (e.g., a clockwise turn on the rotary dial makes the cursor on the center stack move in a downward direction -- or an upward direction for right side driving vehicles). At present, it may prove prudent for motorists to avoid inconsistent center console interactions, as they tend to be associated with higher levels of workload.

Third, are IVIS interactions easier to perform in some vehicles than others? As illustrated in Figure 22, there were surprisingly large differences between vehicles in the overall demand of IVIS interactions. Seven of the 30 vehicles received an overall rating significantly below 1.0 (i.e., a moderate level of overall demand). Eleven of the 30 vehicles received a score that did not differ from the high demand referent (i.e., a high overall demand score). Twelve of the 30 vehicles scored significantly above the high demand referent (i.e., a very high overall demand score). On the whole, vehicles in the latter category tended to have higher levels of demand on cognitive, visual, and subjective measures as well as longer interaction times.

Eighty-three percent of the vehicles with a very high overall demand score offered destination entry for navigation while the vehicle was in motion. However, two out of 12 of the vehicles that offered navigation were not in the very high demand category and two out of 12 of the vehicles in the very high category did not offer navigation entry while the vehicle was in motion. This suggests an association between destination entry to support navigation and overall demand, but the association is not perfect because 17% of the vehicles that offered navigation did so without causing a very high level of demand and 17% of the vehicles that did not offer navigation had a very high level of demand. Nine of the 12 vehicles (75%) in the very high category offered text messaging, whereas 13 of the 18 vehicles (72%) that were not in the very high category offered texting. Apparently, the availability of texting by itself was not sufficient to cause very high levels of demand.

All of the vehicles we evaluated supported auditory vocal IVIS interactions, and 27 out of 30 vehicles supported IVIS interactions using the center stack touch screen. Four of 30 vehicles we tested used the center console for IVIS interactions; one was associated with a moderate demand, one with a high demand, and two were associated with a very high demand. On the whole, there was a modest association between the mode of interaction and the overall workload of the IVIS interactions. However, given that 50% of the vehicles with center console interactions had very high demand, caution should be given to this mode of interaction.

There were unique IVIS features found in only one or two vehicles we tested that were not included in the on-road testing. For example, the Tesla Model S 75 enabled internet usage on the 17-

inch LCD touch screen that allowed the driver to perform a large number of tasks that were not driving-related (e.g., checking and composing Facebook posts). These interactions were clearly distracting and would likely have increased the overall demand rating of the Tesla Model S 75 had this feature been incorporated into the ratings.

The vast majority of the IVIS features and functions in the vehicles we evaluated were unrelated to the task of driving (or, in the case of destination entry to support navigation, could have been performed *before* the vehicle was in motion). Some vehicles supported internet browsing to access social media (e.g., Facebook) while the vehicle was in motion. Others had cumbersome human-machine interfaces with design inconsistencies that lead to high levels of workload. Consumers are often unaware of these IVIS features and how they function (e.g., National Safety Council, 2017). In fact, the IVIS interactions were often associated with high levels of cognitive and visual demand with long interaction times. Our objective assessment indicates that many of these features are just too distracting to be enabled while the vehicle is in motion. Greater consideration should be given to what IVIS features and functions *should* be available to the driver when the vehicle is in motion rather than to what IVIS features and functions *could* be available to motorists.

Theoretical Considerations

It is useful to conceptualize the impairments associated with IVIS interactions stemming from either *structural* or *attentional* interference. The former refers to physical bottlenecks -- the eyes cannot be directed to two locations at the same time and the hands cannot be in two places at once. Structural interference is a physical limitation of the sensory and motor systems. Attentional interference refers to capacity limitations of human attention – a cognitive resource that can be flexibly allocated based on the task goals, motivations, and intentions of an individual (e.g., Kahneman, 1973).

Complex multimodal IVIS interactions involve both structural and attentional interference. Consider the example of a motorist sending a text message. This could begin with pushing a button on the steering wheel to initiate the interaction and continue with uttering a voice-based command (e.g., “send a text message to Jane Doe”), deciding the content of the message and using the voice-to-text features to dictate the message, reviewing the dictated text that is presented on an LCD display located in the center stack, and sending the message. Manually pressing the buttons and looking at the LCD display are examples of structural interference in that they divert the hands and eyes from the task of driving. Other aspects of the texting task involve diverting attention from the task of driving, resulting in various degrees of driver distraction (Regan, Hallett, & Gordon, 2011; Regan & Strayer, 2014).

The current research attempted to provide separate estimates of structural (i.e., visual demand) and attentional (i.e., cognitive demand) sources of interference. However, we acknowledge that few if any tasks are process pure (Jacoby, 1991). Even the SuRT task used in the current study, while placing heavy demands on visual-manual resources (i.e., the eyes and hands), nevertheless placed minimal demands on limited capacity attention. Similarly, N-back tasks such as the one used in the current research have nothing for the driver to touch or see, yet they alter the visual scanning pattern of motorists (for a review, see Strayer & Fisher, 2016). Moreover, the dependent measures are not “pure” either. For example, while the hit rate of the remote DRT is sensitive to eyes off the road (and produces similar patterns to those obtained with eye tracking measures), the inattention blindness literature shows that motorists can look at something and fail to “see” it (e.g., Strayer, Drews, & Johnston, 2003) because attention is diverted by a secondary task that is primarily cognitive in nature.

The separation of structural and attentional interference may be useful for designers to help minimize distraction, so long as there is a realization that both facets of distraction are important to mitigate. Moving from simple button presses to voice commands without a careful analysis of the costs and benefits may have unintended consequences. For example, we found that using voice commands reduced visual demand but at a cost of considerably longer interaction times. In many instances a two-second button press is preferable to a 20-second voice-based interrogatory to perform the same task (see also Kidd et al., 2017).

Finally, the DRT measures are sensitive to changes in the difficulty of both the driving task and concurrent secondary tasks, if any (Turrill & Strayer, In Preparation). Yet the inclusion of the DRT does not appear to alter performance of the driving task (Strayer et al., 2013; see also, Palada et al., 2017). It appears that the DRT draws from the residual capacity that has not been consumed by the primary task of driving or the IVIS secondary tasks. Consequently, the DRT is an exquisite metric for dynamic fluctuations in workload associated with concurrently performing the primary and secondary tasks (e.g., Strayer, Biondi, & Cooper, 2017).

Salvucci (2006, see also Salvucci & Beltowska, 2008) developed a threaded cognition version of ACT-R to predict driving behavior. The model has built in perceptual and motor modules that work in parallel resembling complex human behavior. In addition, there is a cognitive processor that receives all information from the perceptual module and is also in charge of all that goes into the motor module. The "threads" of the cognitive processor operate sequentially. Salvucci (2006) argued that when drivers engage in secondary tasks, the cognitive processor must switch between task threads, resulting in suboptimal driving performance.

Threaded cognition can account for the DRT data discussed above by assuming the processing threads have different priorities such that driving has a higher priority than the IVIS secondary tasks which have a higher priority than responding to the DRT. Interference with driving can occur if an IVIS secondary-task thread "locks out" the processing thread associated with driving. Similarly, driving-related threads will interfere with IVIS secondary-task threads. According to this interpretation, the DRT threads are last to be processed, receiving processing only after critical primary-task (i.e., driving-related) and secondary-task (IVIS-related) threads have been serviced. Consequently, the DRT is sensitive to both primary-task and secondary-task load, but because the task requirements of the DRT are minimal (e.g., press a button if you see a light), any interference (i.e., lock-out) from the DRT thread is minimal.

Limitations and Caveats

One challenge of this research was equating the tasks and modalities of interaction in the different vehicles. Vehicles obviously differ in the features, functions, and human machine interface. Moreover, vehicles often provide more than one way to perform a task and there are often cross-modal interactions wherein a task is initiated using one mode of interaction (e.g., voice commands), and then transitions to another mode of interaction (e.g., touch-screen interactions). We tested each task by mode of interaction available in each vehicle, with some vehicles having missing cells for tasks and/or modes of interaction. If the feature was enabled while driving, it was included in the overall rating of each vehicle, because the collection of features and functions is available to motorists when the vehicle is in motion. Given the association between overall demand ratings and the availability to enter destinations for navigation while driving, vehicles that enable this feature are likely to have increased levels of driver workload. However, there were some vehicles we tested where the navigation task did not lead to high levels of demand, so a system that is designed well and easy to use will not necessarily increase overall workload.

Our research used the experimental method where participants were instructed to perform the IVIS tasks in a counterbalanced order. This method provides an ability to make causal statements regarding different IVIS activities and the workload associated with them. However, in real-world settings, drivers are free to perform the IVIS tasks if, when, and where they so choose. This complicates the relationship between driver workload as measured in experimental studies and crash risk. For example, motorists may attempt to self-regulate their non-driving activities to periods where they perceive the risks to be lower. However, self-regulation depends upon drivers being aware of their performance and adjusting their behavior accordingly, an ability that is often limited by the same factors that caused them to be distracted in the first place (e.g., Sanbonmatsu et al., 2016).

We selected as high-demand referent tasks the N-back (2-back) and SuRT tasks and adopted a 24-second rule for dynamic task interaction time. IVIS interactions (for tasks, modes of interaction, and vehicles) with lower demand than these referent tasks scored well whereas those with higher demand than the referent tasks scored poorly. One may question whether the referents are reasonable. That is, if the referent tasks were too easy (or hard), then the *absolute* ratings would be an over (or under) estimate of the true demand. In fact, the effect size estimates of the N-back and SuRT tasks were very large and of equivalent magnitude (i.e., Cohen's d was 1.423 and 1.519, respectively). Note however, that the *relative* ratings of tasks, modes of interaction, and vehicles should be insensitive to the absolute demand of the referent tasks, so long as they are performed in a consistent fashion in a counterbalanced order across participants. In any event, these two high demand referent tasks have a well-established record for creating high levels of cognitive demand (e.g., Mehler, Reimer, & Dusek, 2011) and visual demand (e.g., Engström & Markkula, 2007; Mattes, Föhl, & Schindhelm, 2007).

The 24-second task interaction referent is derived from the NHTSA visual/manual guidelines (NHTSA, 2013). Video coding of eye glances when participants performed the SuRT task indicated that they took their eyes off the road 50% of the time when performing the SuRT task. Thus, performance of the high visual/manual demand SuRT for 24 seconds, a score of 1.0 in our rating system, matches this threshold. However, this interaction time criterion could be adjusted (e.g., 8.0 seconds for the Japan Automobile Manufacturers Association (2004) guidelines or 20 seconds for the Automobile Alliance guidelines). As above, the change in the task interaction time referent will change the *absolute* ratings; however, the *relative* rank ordering will not change. For example, we tested eight-, 12-, and 15-second task interaction referents and found the same relative ranking of vehicles. What does change with the different referents is the absolute position on the workload scale relative to the red line referent (e.g., all of the vehicles tested score above the high demand referent with either the eight-, 12-, or 15-second limit).

The vehicles we evaluated differed in the number of modes of interaction. Whereas each of the vehicles tested supported auditory-vocal interactions, only four of the vehicles we tested had center console interactions. Consequently, the estimates of demand associated with the latter modality of interaction were more variable than the former (as can be seen in the 95% confidence intervals in Figure 12). In fact, center console interactions tended to be more demanding; however, it is possible that future forms of center console interactions may be less demanding than others. It is also possible that this modality of interaction may continue to be so awkward to use that motorists self-regulate their behavior to avoid using this interface altogether.

Future Directions

The current research examined the demand of IVIS interactions in systems designed by different OEMs. However, a growing trend is to provide access to nomadic systems that support various features and functions. For example, both Apple's CarPlay™ and Google's Android Auto™ are software applications on the iPhone and Android smartphones, respectively, that allow the driver to pair a phone with the vehicle to perform many of the task types and modes of interaction offered by the OEMs. It is currently unknown how these aftermarket systems perform relative to the IVIS systems developed by the OEMs. An extension of the current line of research is presently underway using the protocol developed herein to compare and contrast CarPlay™ and Android Auto™ when they are used in several different vehicles. These apps are often marketed as being easier to use. Do the aftermarket systems vary in different vehicles? How do they compare with each other? How do they compare with the demand of the IVIS systems designed by the OEMs? These are all questions that under study.

Summary

The last decade has seen an extraordinary increase in the digital technology at motorists' fingertips that facilitate IVIS interactions that are unrelated to the task of driving. New vehicles are equipped with (at least) one LCD screen in the center stack that often supports touch-screen interactions with complex menus. All vehicles have some form of voice-command system that allows motorists to push a button and speak to initiate an interaction. Some vehicles include more distinctive configurations (e.g., write pads, rotary dials, gesture controls, heads-up displays, etc.). Surprisingly little is known about how these complex multimodal IVIS interactions impact drivers' workloads. Given the ubiquity of these systems, the current research addressed three interrelated questions concerning this knowledge gap. First, are some tasks more impairing than others? Second, are some modes of interaction more distracting than others? Third, are IVIS interactions easier to perform in some vehicles than others? The answer to each question is yes. Tasks vary in visual, cognitive, and subjective demand as well as the time to perform the actions. Interaction modalities also differ significantly in demand. Moreover, some tasks are less demanding with one mode of interaction (e.g., voice commands to select music) than another (e.g., center console interactions to select music). Finally, vehicles differed considerably in the demand associated with IVIS interactions. Some of the demand stems from the tasks and modes of interaction supported by different OEMs. Other sources of demand were associated with awkward and confusing human machine interfaces. Often, the time to perform an IVIS interaction was excessive. Many of the more complex IVIS features and functions were associated with extreme levels of overall demand.

We suggest that automakers give greater consideration to which IVIS features and functions *should* be available rather than *could* be available when the vehicle is in motion. For example, the NHTSA visual-manual guidelines (2013, p. 116) recommend against in-vehicle electronic systems that allow drivers to perform the following activities when the vehicle is moving:

- Visual-manual text messaging
- Visual-manual internet browsing
- Visual-manual social media browsing
- Visual-manual navigation system destination entry by address
- Visual-manual 10-digit phone dialing
- Displaying more than 30 characters of text unrelated to the task of driving

Our testing found several instances in which drivers could perform the multimodal IVIS interactions listed above while the vehicle was in motion. Notably, vehicles that supported these features when the vehicle was in motion were often associated with the higher demand ratings. Locking out these activities when the vehicle is in motion and shortening the task interaction time are two methods that would reduce the overall demand of the IVIS interactions.

REFERENCES

Angell, L. S., Auflick, J., Austria, P. A., Kochhar, D. S., Tijerina, L., Biever, W., ... & Kiger, S. (2006). *Driver workload metrics task 2 final report* (No. HS-810 635).

Bates, D., Maechler, M., Bolker, B., Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.<doi:10.18637/jss.v067.i01>.

Biondi F., Getty D., Cooper J., Strayer D. (under review). “Sorry, I Didn’t Get That”: Investigating the role of design components on workload and usability of in-vehicle auditory-vocal systems.

Burns, P., Harbluk, J., Foley, J. P., and Angell, L. (2010). The importance of task duration and related measures in assessing the distraction potential of in-vehicle tasks. Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2010), November 11-12, 2010, Pittsburgh, Pennsylvania, USA.

Castro, S., Cooper, J., & Strayer, D. (2016). Validating Two Assessment Strategies for Visual and Cognitive Load in a Simulated Driving Task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 1899-1903.

Castro, S., Strayer, D. L., Matzke, D., & Heathcote, A. (2017). Information-processing dynamics of the Detection Response Task (DRT).

Cooper, J. M., Castro, S. C., & Strayer, D. L. (2016). Extending the Detection Response Task to Simultaneously Measure Cognitive and Visual Task Demands. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 1962-1966.

Davison, A. C., Hinkley, D. V, & Young, G. A. (2003). Recent Developments in Bootstrap Methodology. *Statistical Science*, *18*(2), 141–157.

Domeyer, J. E., Diptiman, T., & Hamada, H. (2014). Using occlusion to measure the effects of the NHTSA participant criteria on driver distraction testing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2205-2209). Sage CA: Los Angeles, CA: SAGE Publications.

Driver Focus-Telematics Working Group. (2006). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems.* Washington, DC: Alliance of Automobile Manufacturers.

Engström, J., & Markkula, G. (2007). Effects of visual and cognitive distraction on lane change test performance. In D. V. McGehee, J. D. Lee, & M. Rizzo (Eds.) *Proceedings of the Fourth International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Published by the Public Policy Center, University of Iowa.*

Fisher, D. L., & Strayer, D. L. (2014). Modeling situation awareness and crash risk. *Annals of Advances in Automotive Medicine, 5, 33-39*.

Getty D., Biondi F., Morgan S., Cooper J., Strayer D. (under review). The effects of voice system design components on driver workload.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323-343.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.

Hothorn, T., Bretz, F., Westfall, P. (2008). Simultaneous Inference in General Parametric Models. Biometrical Journal 50(3), 346--363.

ISO 17488 (2015). Road Vehicles: Transport Information and Control Systems -- Detection-Response Task (DRT) For Assessing Attentional Effects of Cognitive Load in Driving. TC/SC: ISO/TC 22/SC 39.

ISO TS 14198 (2012). Road vehicles – Ergonomic aspects of transport information and control systems – Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems

Ito, H., Atsumi, B., Uno, H., & Akamatsu, M. (2001). Visual distraction while driving; Trends in research and standardization. LATTS research.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.

Japan Automobile Manufacturers Association (2004). *Guideline for In-Vehicle Display Systems, Version 3.0.* Japan Automobile Manufacturers Association. Tokyo, Japan.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Kidd, D. G., Dobres, J., Reagan, I., Mehler, B., & Reimer, B. (2017). Considering visual-manual tasks performed during highway driving in the context of two different sets of guidelines for embedded in-vehicle electronic systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, *47*, 23–33. http://doi.org/10.1016/j.trf.2017.04.002

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives, 7*, 199-204.

Mattes, S., Föhl, U. and Schindhelm, R. (2007). *Empirical comparison of methods for off-line workload measurement*. AIDE Deliverable 2.2.7, EU project IST-1-507674-IP.

McWilliams, T., Reimer, B., Mehler, B., Dobres, J., & McAnulty, H. (2015). Proceedings of the Fourth International Driving Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design. *Driving Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design*, 110–117.

McWilliams, T., Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. F. (2015). Effects of age and smartphone experience on driver behavior during address entry. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15 (pp. 150–153). New York, New York, USA: ACM Press. http://doi.org/10.1145/2799250.2799275

Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J., & McCartt, A. (2015). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics*, *139*(November), 1–24. http://doi.org/10.1080/00140139.2015.1081412

Mehler, B., Reimer, B., & Dusek, J. (2011). MIT AgeLab delayed digit recall cask (n-back) [White paper].

National Safety Council (2017). My Car Does What? Downloaded from the Internet at https://mycardoeswhat.org on June 3, 2017.

National Highway Traffic Safety Administration. (2013). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. *Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT)*.

Palada, H. Strayer, D.L., Neal, A., Ballard, X. & Heathcote A. (2017). A comparison of multitasking performance with and without inclusion of the Detection Response Task (DRT). Paper presented at the Society for Psychonomics Science, Vancouver, British Columbia.

Perez, M., Owens, J., Viita, D., & Angell, L. (2013). Radio tuning effects on visual and driving performance measures - simulator and test track studies. DOT HS 8711 781

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reimer, B., Mehler, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Rumpold, A. (2014). Effects of an "Expert Mode" Voice Command System on Task Performance, Glance Behavior & Driver Physiology. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '14*, 1–9. http://doi.org/10.1145/2667317.2667320

Regan, M. A., Hallett, C. & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis and Prevention*, *43*, 1771-1781.

Regan, M. A., & Strayer, D. L. (2014). Towards an understanding of driver inattention: taxonomy and theory. *Annals of Advances in Automotive Medicine, 58*, 5-13.

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis* (Vol. 4). Champaign, IL: Human kinetics.

Society of Automotive Engineers (2000). Navigation and Route Guidance Function Accessibility While Driving (SAE Recommended Practice J2364, January 20, 2000), Warrendale, PA: Society of Automotive Engineers.

Society of Automotive Engineers (2015). Operational Definitions of Driving Performance Measures and Statistics (Recommended Practice J2944). Society of Automotive Engineers, Warrendale, PA.

Sanbonmatsu, D. M., Strayer, D. L., Biondi, F., Behrends, A. A., & Moore, S. M. (2016). Cell phone use diminishes self-awareness of impaired driving. *Psychonomic Bulletin & Review, 23*, 617-623.

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors*, *48*, 362-380.

Salvucci, D. D. & Beltowska, J. (2008). Effects of memory rehearsal on driver performance: Experiment and theoretical account. *Human Factors, 50,* 834-844.

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis* (Vol. 4). Champaign, IL: Human kinetics.

Schneider, W. & R. M. Shiffrin. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review*, 84, 1-66.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84,127-190.

Shutko, J., and Tijerina, L. (2006). Eye glance behavior and lane exceedences during driver distraction. Presentation given at *Driver Metrics Workshop*, Ottawa, October, 2006. Website http://ppc.uiowa.edu/drivermetricsworkshop/.

Strayer, D. L., Biondi, F., & Cooper, J. M. (2017). Dynamic workload fluctuations in driver/non-driver conversational dyads. In D. V. McGehee, J. D. Lee, & M. Rizzo (Eds.) *Driving Assessment 2017: International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design (pp. x-x). Published by the Public Policy Center, University of Iowa.*

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2016). Talking to your car can drive you to distraction. *Cognitive Research: Principles & Implications, 1, 1-16*. doi: 10.1186/s41235-016-0018-3

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2017). The smartphone and the driver's cognitive workload: A comparison of Apple, Google, and Microsoft's intelligent personal assistants. *Canadian Journal of Experimental Psychology, 71*, 93-110.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile. *AAA Foundation for Traffic Safety.*

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, 9, 23-52.

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Human Factors, 53*, 1300-1324.

Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). Cognitive distraction while multitasking in the automobile. In B. Ross (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory (Vol 54)* (pp. 29-58). San Diego, CA: Elsevier Academic Press.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Tsimhoni, O., Yoo, H., & Green, P. (1999). Effects of visual demand and in-vehicle task complexity on driving and task performance as assessed by visual occlusion. Technical Report UMTRI-99-37.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2 SPEC. ISS.), 167–190. http://doi.org/10.1016/j.trf.2005.04.014

Wickens, C. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. http://doi.org/10.1518/001872008X288394.

Young, R. A., Aryal, B. J., Muresan, M., Ding, X., Oja, S., & Simpson, S. N. (2005). Road-to-lab: Validation of the static load test for predicting on-road driving performance while using advanced in-vehicle information and communication devices. In *Proceedings of the Third*

International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. (pp. 1–15). Iowa City: University of Iowa, Public Policy Center.

Zhang, Y., Angell, L., Pala, S., and Shimonomoto, I., "Bench-Marking Drivers' Visual and Cognitive Demands: A Feasibility Study," *SAE Int. J. Passeng. Cars - Mech. Syst.* 8(2):584-593, 2015, doi:10.4271/2015-01-1389.

APPENDIX 1

A. Audio Entertainment Tasks *all*
- **Radio frequency tuning**
- **iPod contents (songs, artists, albums, genres)**

1. Choose a **jazz song** from the iPod 2. Play **1020 AM** 3. Tune the radio to **98.5 FM** 4. Listen to the song **“99 Red Balloons”** 5. The **band Nirvana** is what you want to hear 6. Change the radio to your favorite **FM station** 7. Turn on the **metal genre** 8. Let’s hear the **song “I’m Gonna Be (500 Miles)”** 9. You want to hear one of your favorite **AM stations** 10. Tune **AM 1540** 11. **89.1** 12. Tune to **1240** 13. iPod play **album “Storyline”**	14. **AM 1160** 15. **Play a song by the artist Eminem** 16. Play the **album “Homesick”** 17. **90.1** 18. You want **Johnny Cash songs** to play 19. Radio **1630** 20. **“Riptide” is a song you want to play** from the iPod 21. Switch the iPod to **artist Louis Armstrong** 22. Play the **alternative genre** 23. Change the **genre** to **reggae** 24. Radio tune to **97.1 FM** 25. You want to hear a song by the **artist Hunter Hayes** 26. Change the music to the **song “Three Little Birds”** 27. Listen to **FM 99.5**

B. Audio Entertainment Tasks *Radio and iPod as source*

- **Radio frequency tuning**
- **iPod source (no content access)**

1. Choose a song from the **iPod**	*12.* **iPod** play
2. Play **1020 AM**	*13.* **AM 1160**
3. Tune the radio to **98.5 FM**	*14.* Play a song via **iPod**
4. Listen to the song on the **iPod**	*15.* **90.1**
5. Change the radio to your favorite **FM station**	*16.* You want **iPod** songs to play
6. Turn on the **iPod music**	*17.* Radio **1630**
7. Let's hear a song on the **iPod**	*18.* You want to play music from the **iPod**
8. You want to hear one of your favorite **AM stations**	*19.* Radio tune to **97.1 FM**
9. Tune **AM 1540**	*20.* You want to hear a song by your favorite artist on the **iPod**
10. **89.1**	*21.* Listen to **FM 99.5**
11. Tune to **1240**	

C. Audio Entertainment Tasks *Radio and iPod songs*

- **Radio frequency tuning**
- **iPod contents (songs only)**
- **Task is complete once media has been loaded.**

1. Choose a **song** from the iPad *2.* Play **1020 AM** *3.* Tune the radio to **98.5 FM** *4.* Listen to the song **"99 Red Balloons"** *5.* The **song "Not Afraid"** is what you want to hear *6.* Change the radio to your favorite **FM station** *7.* Turn on **"Mess Around" the song** *8.* Let's hear the **song "I'm Gonna Be (500 Miles)"** *9.* You want to hear one of your favorite **AM stations** *10.* Tune **AM 1540** *11.* **89.1** *12.* Tune to **1240** *13.* iPod play **song "Storyline"**	*14.* **AM 1160** *15.* Play the **song "Don't Stop Believin'"** *16.* Play the **song "I Can See Clearly Now"** *17.* **90.1** *18.* You want **"If It Means a Lot to You" song** to play *19.* Radio **1630** *20.* **"Riptide" is a song you want to play** from the iPod *21.* Switch the iPod to **song "Let It Be"** *22.* Play the **"Come Together"** *23.* Radio tune to **97.1 FM** *24.* Change the music to the **song "Three Little Birds"** *25.* Listen to **FM 99.5**

D. Audio Entertainment Tasks *iPod only*

- **iPod contents (songs only)**
- **Task is complete once media has been loaded.**

1. Listen to a **jazz song** *2.* Play **the artist A Day to Remember** *3.* Play your **favorite song** *4.* Listen to the song **"99 Red Balloons"** *5.* The **band Nirvana** is what you want to hear *6.* Play a **pop** music internet radio station *7.* Turn on the **metal genre** *8.* Let's hear the **song "I'm Gonna Be (500 Miles)"** *9.* You want to hear one of your favorite **artists** *10.* You want to hear a song by **Adele** 11. "**Somebody Else"** *12.* Tune to **a country** internet radio station *13.* Play the **album "Safe House"**	14. **Pantera** 15. Play a **song** by the **artist Eminem** *16.* Play the **album "Homesick"** *17.* You want to hear the song "**I Can See Clearly Now"** *18.* You want a **Johnny Cash song** to play *19.* **Hip-hop** Radio *20.* **"Riptide" is a song** you want to play *21.* Switch audio to **artist Louis Armstrong** *22.* Play the **alternative genre** *23.* Change the **genre** to **reggae** 24. **"Wagon Wheel"** *25.* You want to hear a song by the **artist Hunter Hayes** *26.* Change the music to the **song "Three Little Birds"** *27.* Listen to **Justin Bieber**

E. Calling & Dialing Tasks *Contacts and Dialing*

- **Participant calls contacts (cell phone, work).**
- **Dials numbers (participant's own phone number+, 801-555-1234).**
- **Task is complete once call has been successfully ended.**

1. **Jack Olsen** would like you to call him on his **cell phone** 2. You need to call **8"OH"1-555-1234** 3. **Willow Brooks** 4. Try to reach **Brad Peterson** 5. Enter **8"ZERO"1-555-1234** 6. You can't find **your phone**. Call it to find it. 7. Ring **Felicity Gomez's office** 8. Enter your **own number** 9. You missed a call from **Oliver Reed** 10. Telephone **8"OH"1-555-1234** 11. **Violet Wheeler** is waiting to hear back from you on her **mobile**	12. Dial your **own number** 13. Give **Phil Potter** a call back at work 14. Give **8"ZERO"1-555-1234** a call 15. Try **Helen Harold** on her **business number** 16. Call your **own phone** 17. **8"OH"1-555-1234** 18. **Bethany Swan, cell phone** 19. Telephone **Jennifer Long** 20. You need to talk to **Yolanda Chavez** 21. Dial **Tanya Henry** 22. Call **Andrew Fink's mobile** back

+RA asked for consent from participant to use participant's phone number.

F. Calling & Dialing Tasks *Contacts only*

- **Participant calls contacts (cell phone, work).**
- **Task is complete once call has been successfully ended.**

1. **Jack Olsen** would like you to call him on his **cell phone**	12. Dial **Tanya Henry**
2. Willow Brooks	13. Call **Andrew Fink's mobile** back
3. Try to reach **Brad Peterson**	14. You need to call **Ian Gavin**
4. Ring **Felicity Gomez's office**	15. Place a call to **Frank Waterfall's office**
5. You missed a call from **Oliver Reed**	16. You can't reach **Francis Baker**. Call them again.
6. **Violet Wheeler** is waiting to hear back from you on her **mobile**	17. You need to reach **Eve Remington**
7. Give **Phil Potter** a call back at **work**	18. Telephone **Daniel Granger**
8. Try **Helen Harold** on her **business number**	19. Dial **Alan Fink's mobile**
9. **Bethany Swan, cell phone**	20. Give **Mia Aston** a call
10. Telephone **Jennifer Long**	21. Call **Nathan Chow** again
11. You need to talk to **Yolanda Chavez**	22. **Oakley James** is waiting for your call

G. SMS *Read Only* Tasks

- **Participant has system read out text messages.**
- **Task complete once message has been selected, not once message is done being read aloud.**

1. Read out the message from **Cam Whitman** 2. Read out the text from **Andy Cameron** 3. What did **Amelia Kidder** send you? 4. **Maggie Carter** just messaged you. 5. What did **Rachel Gatsby** say? 6. Find a message from **Andy Cameron** 7. **Scarlett Miles** sent you a new text 8. Read the text from **Amelia Kidder** 9. What did **Maggie Carter** send you?	10. Read the text from **Cam Whitman** 11. New message from **Lucas Forester** 12. What did **Cam Whitman** send you? 13. Read out the message from **Maggie Carter** 14. What did **Scarlett Miles** send you? 15. What does the text from **Maggie Carter** say? 16. What did **Rachel Gatsby** send you? 17. What does the message from **Scarlett Miles** say? 18. New message from **Andy Cameron**

H. SMS *Read & Send* Tasks

- **Participant reads and responds to texts with system-specific predetermined messages.**
- **Task is complete once message has been sent.**

1. Read out the message from **Cam Whitman**. Please respond.	10. Read and respond to the text from **Cam Whitman**
2. Read and reply to the text from **Andy Cameron**	11. New message from **Lucas Forester**. How do you reply?
3. What did **Amelia Kidder** send you? Send your answer.	12. What did **Cam Whitman** send you? Answer him.
4. **Maggie Carter** just messaged you. What should you send back?	13. Read out the message from **Maggie Carter.** Send your reply.
5. What did **Rachel Gatsby** say? Reply to her.	14. What did **Scarlett Miles** send you? Text her back.
6. Find a message from **Andy Cameron**. Reply.	15. How do you respond to the text from **Maggie Carter**?
7. **Scarlett Miles** sent you a new text. Send something back.	16. You need to read and reply to **Rachel Gatsby's** message.
8. Read the text from **Amelia Kidder** and respond to it.	17. What does the message from **Scarlett Miles** say? Respond.
9. What did **Maggie Carter** send you? Send a text back.	18. Read and then reply to the new message from **Andy Cameron.**

I. SMS *Send Only* Tasks

- **Participant sends a new text message in response to the given scenario to a phone contact.**
- **Task is complete once message has been sent.**

1. Let **Hugo Grant's office** know you're going to be late.
2. **Hunter Bowman** is asking if you want to go to the movies tonight.
3. **Eve Remington** wants to go dancing tonight.
4. **Milly Jung** texted you a funny joke.
5. Text **Kevin Malcome** to ask for directions.
6. Ask **Quinn Brown (cell)** where they are.
7. Tell **Paige Green** you're too busy driving to text right now.
8. Tell **Landon Carter** to text you.
9. **Vince Handcock** texted you a silly dad joke.
10. **Brad Peterson** has big news and is wondering if you can talk right now.
11. **Zoe Ferris** dropped off your favorite cookies at your house.
12. **Isabelle Morales** is wondering where you are.
13. **Natalie Ling** can pick you up from the airport next week.
14. Tell **Jack Olsen** to call you from **work**
15. Tell **Willow Brooks** you're too busy driving to call right now.
16. **Francis Baker** wants to know if they can copy your homework.
17. **Milo Santiago** wants to know why you're not at the restaurant yet.
18. **Gretchen Warner** says she will clean your car for you tonight.

J. Navigation Tasks

- **Participant sets the destination to a point of interest that best fits the task goal.**
- **Participant cancels the route before the task is considered to be complete.**

1. **(Gas)** Fill up at the closest gas station. 2. **(Library)** Your library book is overdue. Let's return it at the closest library. 3. **(Italian restaurant)** You're headed out for some Italian food at nearby restaurant. 4. **(Coffee shop)** Grab yourself a cup of coffee from the closest Starbucks. 5. **(Grocery store)** You need some items from Whole Foods. 6. **(ATM\bank)** You need to get cash from a Wells Fargo bank. 7. **(Mexican Restaurant)** Find a Mexican restaurant nearest you.	8. **(Hospital)** Go visit your friend at the LDS Hospital. 9. **(Chinese restaurant)** You're craving food from Panda Express. 10. **(Movie theater)** You're on your way to see a movie at the nearby theater. 11. **(Hotel/Motel)** Drive to the nearest lodging to stay the night. 12. **(Post office)** You have a package to drop off at the closest Post Office. 13. **(Museum)** Go check out the new exhibit at the Utah Museum of Natural History. 14. **(Shopping center)** Go pick out some new clothes at a nearby shopping mall.

K. Navigation Tasks (with other locations)[+]

- **Participant sets the destination to a point of interest that best fits the task goal.**
- **Participant cancels the route before the task is considered to be complete.**

1. **(Gas)** Fill up at the closest gas station. *2.* **(Bowling)** It's league night at the bowling alley. Don't be late! *3.* **(Restaurant)** You're headed out for some pizza at nearby restaurant. *4.* **(Coffee shop)** Grab yourself a cup of coffee from the closest cafe. *5.* **(Car wash)** Go treat your car to a nice wash and wax. *6.* **(ATM\Bank)** You need to get cash from a Wells Fargo bank. *7.* **(Restaurant)** Enjoy a bagel from Einstein's.	*8.* **(Hospital)** Go visit your friend at the hospital. *9.* **(Golf)** Play 18 holes at a Bonneville Golf Course. *10.* **(Bar)** You're meeting up with your friends at a nearby bar. *11.* **(Hotel/Motel)** Drive to the nearest lodging to stay the night. *12.* **(Police station)** You have to pay a parking ticket at the police station. *13.* **(Museum)** Check out the new rides at Seven Peaks water park. *14.* **(Rest area)** Take a break at Jordanelle rest area.

[+]Use these destinations if vehicle does not support destinations from list J.

L. Navigation Tasks (with other locations)[+]

- **Participant sets the destination to a point of interest that best fits the task goal.**
- **Participant cancels the route before the task is considered to be complete.**

1. **(Gas)** Fill up at the closest gas station. *2.* **(Bookstore)** You are eager to buy a new book. Let's find a bookstore. *3.* **(Breakfast restaurant)** You're headed out for some breakfast at a nearby restaurant. *4.* **(Coffee shop)** Grab yourself a cup of coffee from the closest Starbucks. *5.* **(Pharmacy)** You need some items from the local pharmacy. *6.* **(Camping/ RV parks)** You need to be in nature; find the closest camping spot.	*7.* **(Car wash)** Go treat your car to a nice wash and wax. *8.* **(Bar)** You're meeting up with your friends at the cocktail lounge, Bourbon House. *9.* **(Airport)** You're on your way to pick up a friend at the Salt Lake International Airport. *10.* **(Hotel/Motel)** Drive to the nearest lodging to stay the night. *11.* **(Shopping Center/Mall)** Go on a shopping date at Trolley Square.

[+]Use these destinations of vehicle does not support destinations from list J or K.

M. Audio Entertainment Tasks *Radio and iPod contents*

- **Radio frequency tuning**
- **iPod contents (songs, artists, albums, genres)**

1. Tune the radio to 90.1 FM	13. Tune the radio to 1320 AM
2. Tune the radio to 1230 AM	14. Tune the radio to 107.9 FM
3. Play **Michael Jackson's "Thriller"**	15. Play **Katy Perry's "Rise"**
4. Tune the radio to 94.1 FM	16. Tune the radio to 1490 AM
5. Tune the radio to 530 AM	17. Tune the radio to 103.5
6. Play a **pop genre** song	18. Play a **country genre** song
7. Tune the radio to 98.1 FM	19. Tune the radio to 820 AM
8. Tune the radio to 1160 AM	20. Tune the radio to 96.7 FM
9. Play **The Beatles' "Let It Be"**	21. Play **Adele's "Send My Love"**
10. Tune the radio to 830 AM	22. Tune the radio to 1550 AM
11. Tune the radio to 96.3 FM	23. Tune the radio to 98.1 FM
12. Play an **alternative genre** song	24. Play a **rock genre** song

N. Audio Entertainment Tasks *Radio and iPod Songs*

- **Radio frequency tuning**
- **iPod songs only**

1. Tune the radio to 90.1 FM	11. Tune the radio to 107.9 FM
2. Play the song **“Thriller”**	12. Play the song **“Rise”**
3. Tune the radio to 530 AM	13. Tune the radio to 820 AM
4. Play the song **“Let it Be”**	14. Play the song **“Billie Jean”**
5. Tune the radio to 97.1 FM	15. Tune the radio to 89.1 FM
6. Play the song **“Come Together”**	16. Tune the radio to 1620 AM
7. Tune the radio to 1620 AM	17. Play the song **“Send my Love”**
8. Tune the radio to 96.3 FM	18. Tune the radio to 98.1 FM
9. Play the song **“Don’t Stop Believin’”**	19. Play the song **“H.O.L.Y.”**
10. Play the song **“The Funeral”**	20. Tune the radio to 610 AM

O_1. Audio Entertainment Tasks *Radio, iPod contents and Bluetooth*

- **Radio frequency tuning**
- **iPod contents (songs, artists, albums, genres)**
- **Bluetooth audio source**

1. Tune to 90.1 FM	13. Tune to 1610 AM
2. Change source to Bluetooth Audio	14. Tune to 107.9 FM
3. Play **Michael Jackson's 'Thriller'**	*15.* Change source to **Bluetooth Audio**
4. Tune to 94.1 FM	16. Tune to 1490 AM
5. Tune to 530 AM	*17.* Tune to 103.5 FM
6. Change source to **Bluetooth Audio**	*18.* Play a **country genre** song
7. Tune to 98.1 FM	19. Tune to 820 AM
8. Tune to 1160 AM	*20.* Tune to 96.7 FM
9. Play a **pop genre** song	*21.* Play **Adele's 'Send My Love'**
10. Tune to 830 AM	22. Tune to 1550 AM
11. Tune to 96.3 FM	23. Tune to 98.3 FM
12. Play an **alternative genre** song	*24.* Change source to **Bluetooth Audio**

O_2. Audio Entertainment Tasks *Radio, iPod contents and Bluetooth*

- **Radio frequency tuning to presets only**
- **iPod contents (songs, artists, albums, genres)**
- **Bluetooth audio source**

1. Tune the radio to 90.1 FM	13. Tune the radio to 1620 AM
2. Tune the radio to 1320 AM	14. Tune the radio to 107.9 FM
3. Play **Michael Jackson "Thriller"**	15. Play music via **Bluetooth**
4. Tune the radio to 97.1 FM	16. Tune the radio to 610 AM
5. Tune the radio to 530 AM	17. Tune the radio to 99.5 FM
6. Play music via **Bluetooth**	18. Play a **country genre** song
7. Tune the radio to 98.1 FM	19. Tune the radio to 820 AM
8. Tune the radio to 1160 AM	20. Tune the radio to 96.3 FM
9. Play a **pop genre** song	21. Play **Adele's "Send My Love"**
10. Tune the radio to 820 AM	22. Tune the radio to 1160 AM
11. Tune the radio to 96.3 FM	23. Tune the radio to 98.1 FM
12. Play an **alternative genre** song	24. Play music via **Bluetooth**

P. Audio Entertainment Tasks *Radio Only*
- **Radio frequency tuning**
- **Radio categories**

1. Tune the radio to 90.1 FM	13. Tune the radio to 1320 AM
2. Tune the radio to 1230 AM	14. Tune the radio to 107.9 FM
3. Change Music Type to **nostalgia**	15. Change Music Type to **weather**
4. Tune the radio to 94.1 FM	16. Tune the radio to 1490 AM
5. Tune the radio to 530 AM	17. Change Music Type to **adult hits**
6. Tune the radio to 97.1 FM	18. Tune the radio to 103.5
7. Change Music Type to **R&B**	19. Tune the radio to 820 AM
8. Tune the radio to 1160 AM	20. Tune the radio to 96.7 FM
9. Tune the radio to 101.9 FM	21. Change Music Type to **country**
10. Change Music Type to **college**	22. Tune the radio to 1550 AM
11. Tune the radio to 830 AM	23. Tune the radio to 98.1 FM
12. Tune the radio to 96.3 FM	24. Tune the radio to 610 AM

Q. Audio Entertainment Tasks *Radio and Bluetooth Audio*

- **Radio frequency tuning**
- **Radio categories**

1. Tune the radio to 90.1 FM	13. Tune the radio to 1320 AM
2. Tune the radio to 1230 AM	14. Tune the radio to 107.9 FM
3. Change radio to **satellite**	15. Change radio to **Bluetooth audio**
4. Tune the radio to 94.1 FM	16. Tune the radio to 1490 AM
5. Tune the radio to 530 AM	17. Change radio to **satellite**
6. Tune the radio to 97.1 FM	18. Tune the radio to 103.5
7. Change radio to **Bluetooth audio**	19. Tune the radio to 820 AM
8. Tune the radio to 1160 AM	20. Tune the radio to 96.7 FM
9. Tune the radio to 101.9 FM	21. Change radio to **Bluetooth audio**
10. Change radio to **satellite**	22. Tune the radio to 1550 AM
11. Tune the radio to 830 AM	23. Tune the radio to 98.1 FM
12. Tune the radio to 96.3 FM	24. Tune the radio to 610 AM

R. Calling & Dialing Tasks *Contacts and Dialing*

- **Participant calls contacts (cell phone, work).**
- **Dials numbers (participant's own phone number+, 801-555-1234).**
- **Task is complete once call has been successfully ended.**

1. Call **George Hudson** at **work**	11. Dial your **own phone #**
2. Dial your **own phone #**	12. Call **Wendy Darling**
3. Call **Matt Plumb's mobile**	13. Call **Jessica Day**
4. Dial **801-555-1234**	14. Dial **801-555-1234**
5. Call **David Jones** at **work**	15. Call **Ethan Hawke**
6. Dial your **own phone #**	16. Dial your **own phone #**
7. Call **Zane Thompson**	17. Call **Lisa Hamilton**
8. Dial **801-555-1234**	**18.** Dial **801-555-1234**
9. Call **Randall Jenkins' mobile**	19. Call **Brittany Sanders' mobile**
10. Call **William Dunn** at **work**	20. Dial your **own phone #**

+RA asked for consent from participant to use participant's phone number.

S. Calling & Dialing Tasks *Contacts only*

- **Participant calls contacts (cell phone, work).**
- **Task is complete once call has been successfully ended.**

1. Call **William Dunn** at **work**	11. Call **Anna Pearl**
2. Call **Phil Dunphee**	12. Call **Randall Jenkins** at **work**
3. Call **Jessica Day**	13. Call **Ethan Hawke**
4. Call **Matt Plumb's mobile**	14. Call **Brittany Sanders** at **work**
5. Call **Ethan Hawke**	15. Call **David Jones' mobile**
6. Call **Brittany Sanders' mobile**	16. Call **Phil Dunphee**
7. Call **Lisa Hamilton**	**17.** Call **Zane Thompson**
8. Call **Wendy Darling**	18. Call **Randall Jenkin's mobile**
9. Call **George Hudson** at **work**	19. Call **Anna Pearl**
10. Call **David Jones' Mobile**	20. Call **William Dunn** at **work**

T. Calling & Dialing Tasks *Favorite Contacts*

- **Participant calls contacts that have been stored as favorites if system does not allow access to the phonebook while driving.**
- **Task is complete once call has been successfully ended.**

1. Call **William Dunn**	11. Call **George Hudson**
2. Call **Lisa Hamilton**	12. Call **Lisa Hamilton**
3. Call **David Jones**	13. Call **David Jones**
4. Call **Zane Thompson**	14. Call **Zane Thompson**
5. Call **Randall Jenkins**	15. Call **Randall Jenkins**
6. Call **Matt Plumb**	16. Call **Matt Plumb**
7. Call **Jessica Day**	17. Call **Jessica Day**
8. Call **Wendy Darling**	18. Call **Wendy Darling**
9. Call **Zane Thompson**	19. Call **Zane Thompson**
10. Call **Ethan Hawke**	20. Call **Ethan Hawke**

U. SMS *Send Only* Tasks

- **Participant sends a new text message. System only replies to most recent message in the inbox.**
- **Task is complete once message has been sent.**

1. Send “**Yes**”	11. Send “**Be there in 20 minutes**”
2. Send “**Where are you?**”	12. Send “**Why?**”
3. Send “**I’m stuck in traffic**”	13. Send “**I love you**”
4. Send “**No**”	14. Send “**Call you later. I’m driving.**”
5. Send “**Can’t talk right now, I’m driving**”	15. Send “**I need more directions. Can you call me?**”
6. Send “**Be there in 10 minutes**”	16. Send “**Too funny.**”
7. Send “**Can’t wait to see you**”	17. Send “**Call me. I’m driving.**”
8. Send “**Too funny**”	18. Send “**Yes**”
9. Send “**Thanks**”	19. Send “**No**”
10. Send “**Call me. I’m driving.**”	20. Send “**I’m stuck in traffic.**”

V. SMS *Send New Message* Tasks

- **Participant sends a new text message to a phone contact.**
- **Task is complete once message has been sent.**

1. Send "**Text me the address**" to **Jessica Day** 2. Send "**I'll arrive soon**" to **Phil Dunphee** 3. Send "**Stuck in traffic**" to **David Jones' cell** 4. Send "**LOL**" to **Valentine Wiggin** 5. Send "**Where are you?**" to **Wendy Darling** 6. Send "**Thanks**" to **William Dunn** at **work** 7. Send "**Yes**" to **Brittany Sanders' cell**	8. Send "**No**" to **Ethan Hawke** 9. Send "**On my way**" to **Randall Jenkins** at **work** 10. Send "**Call me**" to **Jessica Day** 11. Send "**LOL**" **David Jones** at **work** 12. Send "**When?**" to **Wendy Darling** 13. Send "**Yes**" to **Anna Pearl** 14. Send "**Where?**" to **George Hudson's mobile** 15. Send "**Text me the address**" to **Jessica Day**

W. SMS *Read Only* Tasks

- **Participant has system read out text messages.**
- **Task complete once message has been selected, not once message is done being read aloud.**

1. Read text from **Maddie McCarty**	11. Read text from **Andrea Campos**
2. Read text from **Sydney Mills**	12. Read text from **Kelly Mckenzie**
3. Read text from **Camille Wheatley**	13. Read text from **Maddie McCarty**
4. Read text from **Kelly Mckenzie**	14. Read text from **Camille Wheatley**
5. Read text from **Andrea Campos**	15. Read text from **Rachelle Gatsby**
6. Read text from **Rachelle Gatsby**	16. Read text from **Sydney Mills**
7. Read text from **Camille Wheatley**	17. Read text from **Maddie McCarty**
8. Read text from **Sydney Mills**	18. Read text from **Andrea Campos**
9. Read text from **Andrea Campos**	19. Read text from **Camille Wheatley**
10. Read text from **Rachelle Gatsby**	20. Read text from **Kelly Mckenzie**

X_1. Navigation Tasks

- **Participant sets the destination to a point of interest that best fits the task goal.**
- **Participant cancels the route before the task is considered to be complete.**

1. Find the nearest **Dunkin' Donuts**	11. Find the nearest **gas station**
2. Find the nearest **Mexican**	12. Find the nearest **museum**
3. Find the nearest **hospital**	13. Find the nearest **post office**
4. Find the nearest **Chase Bank**	14. Find the nearest **Hilton hotel**
5. Find the nearest **Starbucks**	15. Find the nearest **breakfast**
6. Find the nearest **gas station**	16. Find the nearest **barbecue**
7. Find the nearest **hotel**	17. Find the nearest **Whole Foods**
8. Find the nearest **coffee shop**	18. Find the nearest **Japanese**
9. Find the nearest **Wells Fargo**	19. Find the nearest **coffee shop**
10. Find the nearest **library**	20. Find the nearest **Best Western**

X_2. Navigation Tasks

- **Participant sets the destination to a point of interest that best fits the task goal.**
- **Participant cancels the route before the task is considered to be complete.**

1. Navigate to the closest **hospital**	11. Navigate to the closest **city center**
2. Navigate to the closest **restaurant**	12. Navigate to the closest **ski resort**
3. Navigate to the closest **hotel and motel**	13. Navigate to the closest **golf course**
4. Navigate to the closest **gas station**	14. Navigate to the closest **embassy**
5. Navigate to the closest **shopping center**	15. Navigate to the closest **campground**
6. Navigate to the closest **ATM**	16. Navigate to the closest **rest area**
7. Navigate to the closest **theater**	17. Navigate to the closest **business facility**
8. Navigate to the closest **museum**	18. Navigate to the closest **train station**
9. Navigate to the closest **coffee shop**	19. Navigate to the closest **tourist attraction**
10. Navigate to the closest **police station**	20. Navigate to the closest **pharmacy**

Use these destinations of vehicle does not support destinations from list X_1.

Y. Audi Q7 3.0T Premium Plus Audio Entertainment Tasks

- **Use the draw pad to select music**
- **Use the alphabet ribbon to search for music**
- **Use the rotary wheel to search for music**

1. Using the **draw pad**, tune the radio to 97.1 FM 2. Play the song “Send My Love To My New Lover” using the **draw pad** 3. Play the artist the Beatles using the **alphabet ribbon** 4. Using the **frequency list and presets**, tune the radio to 1160 AM 5. Play “25” using the **album menu** 6. Play music via Bluetooth 7. Using the **alphabet ribbon**, tune the radio to FM 100.3 FM 8. Using the **draw pad**, play an Alternative genre song 9. Play the artist Katy Perry using the **artist menu** 10. Using the **frequency list and presets**, tune the radio 1240 AM	11. Play the song “Let It Be” using the **alphabet ribbon** 12. Play the artist “Justin Bieber” using the **draw pad** 13. Play music via Bluetooth 14. Using the **frequency list and presets**, tune the radio to 96.3 FM 15. Play the song “Don’t Stop Believin’” using the **alphabet ribbon** 16. Using the **draw pad,** tune the radio to AM 1490 17. Play the artist Michael Jackson using the **artist menu** 18. Play a country genre song using the **genre menu**

Z. Audi Q7 3.0T Premium Plus Calling & Dialing Tasks

- **Participant dials numbers using:**
  - **Alphabet ribbon**
  - **Draw pad**
- **Participant calls contacts (work or cell phone)using:**
  - **Alphabet ribbon**
  - **Draw pad**
  - **Rotary wheel**
- **Task is complete once call has been successfully ended.**

1. **Directory** Call John Smith at work 2. **Draw pad** Dial your own phone # 3. **Directory** Call Matt Plumb's cell phone 4. **Alphabet ribbon** Dial 801-555-1234 5. **Directory** Call William Dunn at work 6. **Draw pad** Dial your own phone # 7. **Draw pad** Call Opal Woods 8. **Alphabet ribbon** Dial 801-555-1234 9. **Alphabet ribbon** Call Frank Fontain at work 10. **Directory** Call Brittany Sanders' cell phone	11. **Alphabet ribbon** Dial your own phone # 12. **Draw pad** Call Helen Harris 13. **Alphabet ribbon** Call Jessica Day 14. **Draw pad** Dial 801-555-1234 15. **Directory** Call Ethan Hawke 16. **Draw pad** Dial your own phone # 17. **Draw pad** Call Lisa Hamilton 18. **Alphabet ribbon** Dial 801-555-1234 19. **Directory** Call Brittany Sanders' Cell Phone 20. **Draw pad** Dial your own phone #

AA. Audi Q7 3.0T Premium Plus *Send New Message* Tasks

- **Participant sends a new text message to a phone contact using:**
  - **Draw pad**
  - **Alphabet ribbon**
- **Task is complete once message has been sent.**

1. **Draw pad** Send "Please Call Back!" to Carly Duncan 2. **Alphabet ribbon** Send "No" to Albert Fink 3. **Draw pad** Send "Congratulations!" to Jessica Day 4. **Alphabet ribbon** Send "I'm in the car, and I'll be late because of traffic" to Eleanor Lamb 5. **Alphabet ribbon** Send "Conference Cancelled" to William Dunn 6. **Draw pad** Send "See you later" Andrew Ryan 7. **Alphabet ribbon** Send "When & where shall we meet?" to David Jones' cell 8. **Draw pad** Send "Ok, thanks for the info" to Ana Pearl 9. **Draw pad** Send "I'm in the car, I'll call later" to Lisa Hamilton 10. **Alphabet ribbon** Send "Ok" to Opal Woods	11. **Alphabet ribbon** Send "Please Call Back!" to Brittany Sanders' Work 12. **Draw pad** Send "No" to Quin Brown 13. **Alphabet ribbon** Send "Congratulations!" to Helen Harris 14. **Draw pad** Send "I'm in the car, and I'll be late because of traffic" to Robert Lutece 15. **Alphabet ribbon** Send "Conference Cancelled" to Ethan Hawke 16. **Draw pad** Send "See you later" to David Jones' work 17. **Draw pad** Send "When & where shall we meet?" to Matt Plumb's cell 18. **Alphabet ribbon** Send "OK, thanks for the info." to Phil Dunphee 19. **Draw pad** Send "I'm in the car, I'll call later" to Helen Harris 20. **Alphabet ribbon** Send "OK" to Zane Thompson

BB. Audi Q7 3.0T Premium Plus Navigation Tasks

- **Participant sets the destination to a point of interest that best fits the task goal using:**
  - **Rotary wheel**
  - **Draw pad**
  - **Alphabet ribbon**
- **Participant cancels the route before the task is considered to be complete.**

1. **Categories menu** Bank	11. **Categories menu** Movie theater
2. **Draw pad** Zuppa's Cafe	12. **Alphabet ribbon** Whole Foods
3. **Alphabet ribbon** Gas station	13. **Draw pad** Post Office
4. **Categories menu** Hotel	14. **Categories menu** Shopping center
5. **Draw pad** Museum	15. **Alphabet ribbon** Hospital
6. **Alphabet ribbon** ATM	16. **Draw pad** Rest Area
7. **Categories menu** Greek restaurant	17. **Categories menu** Winery
8. **Draw pad** NoBrow Coffee	18. **Alphabet ribbon** Pizza
9. **Alphabet ribbon** Fast food	19. **Draw pad** Apple Store
10. **Draw pad** Library	20. **Categories menu** Brewery

APPENDIX 2

Examples of how the statistical contrasts were performed using the lmer function in the lme4 package for R for the dependent measure of cognitive demand. A similar syntax was used for visual demand, subjective demand, interaction time, and overall demand. The number of vehicles driven by a participant was entered as a fixed effect while Participant, Vehicle, Modality, and Task Type were entered as random effects. In each case, p-values were obtained by likelihood ratio tests comparing the simple fixed effects model to the model without the effect in question.

Task Type Analysis

Full Model = lmer(CogDemand ~ Task + Num_Vehicle_Driven + (1|Participant) + (1|Vehicle) + (1|Modality), data)

Partial_Model = lmer(CogDemand ~ Num_Vehicle_Driven + (1|Participant) + (1|Vehicle) + (1|Modality), data)

anova(Full_Model, Partial_Model)

Modality Analysis

Full Model = lmer(CogDemand ~ Modality + Num_Vehicle_Driven + (1|Participant) + (1|Vehicle) + (1|Task), data)

Partial_Model = lmer(CogDemand ~ Num_Vehicle_Driven + (1|Participant) + (1|Vehicle) + (1|Task), data)

anova(Full_Model, Partial_Model)

Task Type by Modality Analysis

Full Model = lmer(CogDemand ~ Task*Modality + Num_Vehicle_Driven + (1|Participant) + (1|Vehicle), data)

Partial_Model = lmer(CogDemand ~ Task + Modality + Num_Vehicle_Driven + (1|Participant) + (1|Vehicle), data)

anova(Full_Model, Partial_Model)

Vehicle Analysis

Full Model = lmer(CogDemand ~ Vehicle + Num_Vehicle_Driven + (1|Participant) + (1|Modality) + (1|Task), data)

Partial_Model = lmer(CogDemand ~ Num_Vehicle_Driven + (1|Participant) + (1|Modality) + (1|Task), data)

anova(Full_Model, Partial_Model)

APPENDIX 3

Expert Evaluators and Evaluation

Three human factors experts are assigned to the evaluation of each vehicle. Experts hold bachelor degrees in experimental psychology, have received extensive training in human factors and ergonomics practices, and work under the supervision of professionals with extended experience in automotive human factors.

The evaluation process consists of three phases. First, evaluators drive each vehicle extensively in order to become familiar with its infotainment functions and human-machine interfaces. Each vehicle is driven in different weather and lighting conditions to identify specific issues that may cause the interaction with the system to be particularly distracting or troublesome (e.g., display glare with low environmental light). After this first familiarization phase, evaluators conduct the formal evaluation of the in-vehicle infotainment system and human-machine interface using the human factors checklist described below. This evaluation phase has a duration of six to 12 hours in which evaluators closely examine separate aspects of the in-vehicle infotainment systems and rate them against the 15 scales in the checklist. A third, examination phase (two to four hours) is conducted for evaluators to compare the information collected inside the vehicle and control for possible inaccuracies. This phase is of importance to ascertain the accuracy of the information in the final report.

Checklist Development

The checklist has a number of subsections for conducting: (1) an overall evaluation of the design structure and layout of the entire human-machine interface of the in-vehicle infotainment system, (2) evaluation of specific infotainment functions (calling, text messaging, navigation, audio entertainment) accessible via available interfaces (touch screen, voice recognition, instrument cluster display, etc.). The checklist was developed upon industry standards and human factors design guidelines. Nielsen's Ten Usability Heuristics for User Interface Design [1] were used as broad design principles to guide in the evaluation of infotainment functions and vehicle interfaces. Five additional design principles were developed and added to the original list of ten Nielsen's heuristics to facilitate the assessment of automotive-specific design and interaction elements. Design guidelines and automotive industry standards were adopted to corroborate the resulting checklist of 15 principles. Design guidelines considered in the development of the checklist include: NHTSA's Human Factors Design Guidance for Driver-Vehicle Interfaces [2], the TRL Checklist for the assessment of in-vehicle infotainment systems [3], the Automobile Alliance's Procedures for the design of in-vehicle information systems [4], the European Commission's Recommendation on safe and efficient in-vehicle information and communication systems [5].

Checklist

Each design principle is followed by its description and selected, automotive-specific references. For each principle, evaluators rate each infotainment function and human-machine interface using a 7-point Likert scale with 1 being deplorable, 4 being average, and 7 being superb.

1- Visibility of the System Status

   The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

   Reference: ISO 15008 [6], ISO 15005 [7]

2- Match between system and the real world

The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

Reference: ISO 15007-1 [8], ISO 15007-2 [9], SAE J2830 [10]

3- User control and freedom

Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

Reference: ESoP (5)

4- Consistency and standards

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

Reference: SAE J2830 [10], ISO 2575 [11]

5- Error prevention

Even better than good error messages is a careful design that prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

Reference: NHTSA [2], Alliance [4]

6- Recognition rather than recall

Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

Reference: NHTSA [2]

7- Flexibility and efficiency of use

Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

Reference: Nielsen [1]

8- Aesthetic and minimalist design

Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

Reference: Federal Highway [12]

9- Help users recognize, diagnose, and recover from errors

Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

Reference: TRL [3]

10- Help and documentation

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focus on the user's task, list concrete steps to be carried out, and not be too large.

Reference: TRL [3], NHTSA [13]

11- Pace of the Task (Internal Locus of Control)

User control of the pace of the task can reduce workload. A system that demands timed/constant input will demand greater attention.

Reference: ISO 15005 [7], TRL [3], ESoP [5]

12- Time on Task (ToT)

The duration of completing a task is often comorbid with efficiency issues and can increase workload.

Reference: NHTSA [13]

13- Reach (Fitt's Law)

The physical ability to reach for relevant buttons or provide touch-screen input based on distance and size of the target button.

Reference: TRL [3], ISO 3958 [14], SAE J 287 [15]

14- Design Errors

Ergonomic and environmental factors that interfere with the system's functionality (e.g., sunlight preventing screen visibility; fan speed interfering with microphone receptivity).

Reference: ISO 15008 [6], TRL [3], ESoP [5], NHTSA [2

15- Safety Concerns

Facets of the system that are immediately apparent (e.g., what aspects are locked out while driving? Dialing numbers, typing out text, responsiveness of touch screens, size of buttons, etc.).

Reference: NHTSA [13]

An online version of the checklist developed using Google Forms is available at the following webpage:
https://drive.google.com/open?id=1pewFvgXbIbhzDcUjCD3NCZn3p9S6wnSrqgZY3an1Glo

REFERENCES

[1] Nielsen, J. (1995). 10 usability heuristics for user interface design. Nielsen Norman Group, 1(1).

[2] Campbell, J. L., Brown. J. L., Graving, J. S., Richard, C. M., Lichty, M. G., Sanquist, T., … & Morgan, J. L.. (2016, December). Human factors design guidance for driver-vehicle interfaces (Report No. DOT HS 812 360). Washington, DC: National Highway Traffic Safety Administration.

[3] Stevens, A., & Cynk, S. (2011). Checklist for the assessment of in-vehicle information systems (Vol. 1, No. 1, pp. 1-35). Transportation Research Laboratory

[4] Alliance of Automobile Manufacturers. (2006). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In Vehicle Information and Communication Systems, 89. Retrieved from http://www.umich.edu/~driving/publications/PGCRCChapter24DRAFT.pdf%5Cn%5Cn

[5] Commission of the European Communities. (2008). ESoP - European Statement of Principles on human-machine interface. Official Journal of the European Union.

[6] ISO (2009). ISO 15008 Road vehicles -- Ergonomic aspects of transport information and control systems -- Specifications and test procedures for in-vehicle visual presentation. 2nd Edition.

[7] ISO (2002). ISO 15005 Road vehicles -- Ergonomic aspects of transport information and control systems -- Dialogue management principles and compliance procedures. 1st Edition.

[8] ISO (2002). ISO 15007-1 Road vehicles -- Measurement of driver visual behavior with respect to transport information and control systems -- Part 1: Definitions and parameters. 1st Edition

[9] ISO (2001). ISO TS 15007-2 Road vehicles -- Measurement of driver visual behavior with respect to transport information and control systems -- Part 2: Equipment and procedures. 1st Edition.

[10] SAE (2008). SAE J2830. Process for comprehension testing of in-vehicle icons. Warrendale, PA: SAE International.

[11] ISO (2010). ISO 2575. Road vehicles—Symbols for controls, indicators and telltales. Geneva: International Organization for Standardization.

[12] Campbell, J. L., Richman, J. B., Carney, C., & Lee, J. D. (2004). In-vehicle display icons and other information elements, Volume I: Guidelines (Report No. FHWA-RD-03-065). Washington, DC: Federal Highway Administration

[13] NHTSA. (2013). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Vol. 77). Retrieved from http://www.distraction.gov/download/distracted_guidelines-FR_04232013.pdf

[14] ISO (1996). ISO 3958 Road Vehicles, Passenger cars - Driver hand control reach. 2nd Edition

[15] SAE (2007). SAE J287 Driver hand control reach.

Table 1. A listing of the IVIS tasks and modes of interaction tested in each vehicle. Specific task lists were developed to test the different vehicles because they supported different combinations features and functions. In the table, the letter combinations (e.g., A-BB) refer to specific task set instructions that are described in detail in Appendix 1. Column headers refer to the different tasks by modality combinations. AE CS refers to audio entertainment performed using the center stack. AE AV refers to audio entertainment performed using the auditory vocal mode of interaction. AE CC refers to audio entertainment using the center console. CD CS refers to calling and dialing performed using the center stack. CD AV refers to calling and dialing performed using the auditory vocal mode of interaction. CD CC refers to calling and dialing using the center console. TXT CS refers to text messaging performed using the center stack. TXT AV refers to text messaging performed using the auditory vocal mode of interaction. TXT CC refers to text messaging using the center console. NAV CS refers to navigation performed using the center stack. NAV AV refers to navigation performed using the auditory vocal mode of interaction. NAV CC refers to navigation using the center console.

Vehicle	Condition											
	AE CS	AE AV	AE CC	CD CS	CD AV	CD CC	TXT CS	TXT AV	TXT CC	NAV CS	NAV AV	NAV CC
Audi Q7 3.0T Premium Plus		O_1	Y		R	Z		W	AA		X_1	BB
Cadillac XT5 Luxury	M	M		R	R							
Chevrolet Equinox LT	M	M		S	R		U					
Chevrolet Traverse LT	M				R							
Chrysler 300C	A	A		E	E		G	H		J	J	
Dodge Durango GT	A	A		E	E		G	H		J	J	
Dodge Ram 1500 Express	M	N		R	R			V				
Ford F250 XLT	O_1	M		T	R		W	W				
Ford Fusion Titanium	M	M		R	R		U	U				
Ford Mustang GT Premium Convertible	A	A		E	E		G	I		J	J	
GMC Yukon SLT	C, A*	A		E	E		H			L	J	
Honda Civic Touring	A	A		F	E		H			J	J	
Honda Ridgeline RTL-E	A	A		F	E		H			J	J	
Hyundai Santa Fe Sport	M	Q		R	R							
Hyundai Sonata Base	A	B		F	E							
Infiniti Q50 3.0T Premium	O_1	O_1	O_2	T	R	T	U	V	U			
Jeep Compass Sport	P	Q			R			V				
Jeep Grand Cherokee Limited	O_1	N		R	R		W	V		X_1	X_1	
Kia Sorento LX	M	Q		R	R							
Lincoln MKC Premiere	O_1	O_1		R	R		W	U				
Mazda3 Touring		M	M		R	R		U	U			
Nissan Armada SV	A			F	E					K	K	
Nissan Maxima SV	M	M	O_2	T	R	T	U	V	U	X_2	X_2	X_2
Subaru Crosstrek Premium	A	B		F	E							
Tesla Model S 75	A	D		E	E					J	J	
Toyota Camry SE	M			S	R		U	V				
Toyota Corolla SE	O_1			S	R		U	V				
Toyota RAV4 XLE	O_1	O_1		S	R		U	V				
Toyota Sienna XLE	O_1	O_1		S	R		U	V				

Volvo XC60 T5 Inscription	**M**	**M**		**R**	**R**		**W**	**W**		**X_2**	**X_2**	
	*Manual (C), Touchscreen (A)											

Table 2. A listing of IVIS tasks performed using the different modes of interaction. Note that the vehicles tested support different combinations of tasks and modes of interaction (see Table 1 for details).

	Center Stack	**Auditory Vocal**	**Center Console**
Audio Entertainment	a) Tune the radio to <*station #*> b) Play <*artist name/song title/genre type*> c) Change the audio source to... <*iPod/AM/FM/XM/Bluetooth*>	a) Tune the radio to <*station #*> b) Play <*artist name/song title/genre type*> c) Change the audio source to... <*iPod/AM/FM/XM/Bluetooth*>	a) Tune the radio to <*station #*> b) Play <*artist name/song title/genre type*> c) Change the audio source to… <*iPod/AM/FM/XM/Bluetooth*>
Calling and Dialing	a) Call <*Contact name at work/on mobile*> b) Dial <*participant's own phone number* c) Dial <*801-555-1234*>	a) Call <*Contact name at work/on mobile*> b) Dial <*participant's own phone number*> c) Dial <*801-555-1234*>	a) Call <*Contact name at work/on mobile*> b) Dial <*participant's own phone number*> c) Dial <*801-555-1234*>
Text Messaging	a) Reply to a text message in the inbox with <*a predetermined message*> b) NA c) Listen to a text message in the inbox from <*Contact Name*>	a) Reply to a text message in the inbox with... <*a predetermined message*> b) Send a new text message to <*Contact Name*> that says <*a pre-determined message*> c) Listen to a text message <*text message*>	a) Reply to a text message in the inbox with <*a predetermined message*> b) Send a new text message to <*Contact Name*> that says <*a pre-determined message*> c) NA
Navigation	a) Navigate to <*point of interest*>.	a) Navigate to <*point of interest*>.	a) Navigate to <*point of interest*>.