AGELAB
Life Tomorrow

MIT Project on Assessing the Demands of Voice Based In-Vehicle Interfaces: Study 2

Technical Report 2014-2

Further Evaluation of the Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Replication and a Consideration of the Significance of Training Method

July 9, 2014

Bruce Mehler, Bryan Reimer, Jonathan Dobres, Hale McAnulty, Alea Mehler, Daniel Munger, & Joseph F. Coughlin

Abstract - This report assesses the extent to which key findings from our initial on-road study (Reimer, Mehler, Dobres & Coughlin, 2013) on driver interaction with a production version, in-vehicle voice command system replicate, as well as considering whether two differing approaches to introducing drivers to the driver vehicle interface (DVI) impact their pattern of interaction, including driving behavior. An analysis sample of 64 participants, equally balanced by gender across the four age groupings (18-24, 25-39, 40-54, and 55+) specified in the NHTSA (2013) visual-manual guidelines for DVI assessment, was evaluated during manual radio tuning, voice-command assisted radio tuning, and voice-command assisted navigation system interaction consisting of full destination address entry and route cancelation under actual highway driving conditions. The MIT AgeLab auditory presentation / vocal response n-back cognitive demand reference task was also presented. No statistically significant main effects of self-guided vs. structured training condition were found across the tasks as a whole, although clear advantages were evident in tasks requiring memorization of complex command syntax. The basic pattern of results seen in the first study (considering self-reported workload, physiological arousal, driving performance metrics, and glance behavior) largely replicated. Voice recognition was again found to be fairly robust with only 3 out of more than 80 participants unable to participate due to voice recognition issues. For the radio tuning reference task, the voice-command method was associated with lower workload (self-report, heart rate, skin conductance level (SCL)), lower mean glance durations, and a markedly lower percentage of long duration glances than the visual-manual interface. Apparent cognitive processing demand / workload as assessed through heart rate and SCL for the DVI tasks studied fell below the level of the 1-back cognitive reference task. Some voice-command

involved tasks, particularly full destination address entry, were associated with a high degree of total eyes off-road time (TEORT). Cognitive demands become more apparent when drivers had difficulty completing activities. Assessment of the "voice" interface in these studies illustrates that in modern DVI's, attentional draws can be highly multimodal (combinations of visual, manual, auditory, vocal, haptic, etc.). Thus, visual as well as other potential demand sources need to be included in the assessment of voice interfaces.

Note: *This document significantly extends and supersedes an earlier project report (2013-20) dated November 30, 2013.*

Lead Project Contact:

Bryan Reimer, Ph.D.
Phone (617) 452-2177
reimer@mit.edu

REPORT DOCUMENTATION PAGE

RECEIVING ORGANIZATION	REPORT DATE	REPORT TYPE + DATES COVERED
Toyota Collaborative Safety Research Center	July 9, 2014	Technical Report

TITLE AND SUBTITLE Further Evaluation of the Effects of a Production Level “Voice-Command” Interface on Driver Behavior: Replication and a Consideration of the Significance of Training Method	FUNDING NUMBERS
AUTHOR(S) Bruce Mehler, Bryan Reimer, Jonathan Dobres, Hale McAnulty, Alea Mehler, Daniel Munger, & Joseph F. Coughlin	
PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) MIT AgeLab Massachusetts Institute of Technology 77 Massachusetts Avenue, E40-279 Cambridge, MA 02139 USA	PERFORMING ORGANIZATION REPORT NUMBER Technical Report 2014-2
SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Toyota Collaborative Safety Research Center Collaborative Safety Research Center Toyota Technical Center USA, Inc. 1555 Woodridge Ave Ann Arbor, MI 48105	SPONSORING/MONITORING AGENCY REPORT NUMBER
SUPPLEMENTARY NOTES This document significantly extends and supersedes an earlier project report, *A Consideration of the Significance of Type of Introductory Training in the Evaluation of the Effects of a Production Level “Voice-Command” Interface on Driver Behavior* (2013-20), submitted to CSRC November 30, 2013.	
DISTRIBUTION/AVAILABILITY STATEMENT	DISTRIBUTION CLASSIFICATION
ABSTRACT (See cover page.)	
SUBJECT TERMS Voice-command, visual-manual demand, TEORT, driver-vehicle interface, DVI, HMI, driver distraction, attention management, workload, cognitive demand, guidelines, glance behavior, heart rate, skin conductance, driving performance.	NUMBER OF PAGES 191

Glossary of Terms (or Acronyms)

ANOVA	analysis of variance
CAMP	Crash Avoidance Metrics Partnership
CAN	controller area network
CSRC	Toyota Collaborative Safety Research Center
DVI	driver vehicle interface
EORT	eyes-off-road time – note that in this metric, all glances away from the forward roadway, including glances to the rear or side mirrors, are counted as off-the-forward-roadway.
GTD	glance to device
M	mean
MIT	Massachusetts Institute of Technology
MoCA	Montreal Cognitive Assessment battery
NHTSA	National Highway Transportation Safety Administration
OR	orienting response
SCL	skin conductance level
SD	standard deviation
Study 1	First study in an ongoing series considering drivers' interactions with voice-command systems – see Reimer, Mehler, Dobres and Coughlin (2013)
Study 2	The study presented in this report
SWR	steering wheel reversals
The Alliance	Alliance of Automotive Manufacturers
TEORT	total eyes-off-road time - the sum of all glances off the forward roadway during a specific period (see EORT)

Further Evaluation of the Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Replication and a Consideration of the Significance of Training Method

Contents

Introduction

Background

Drivers continue to demand increased connectivity and more advanced entertainment options while driving. While automotive manufacturers strive to provide drivers with convenient, safe, easy-to-access information to meet this growing demand, there remains no well-established method for optimally achieving this goal. Over the past several years, there has been a shift in automotive driver-vehicle interfaces (DVIs) from purely visual-manual interactions to include options for voice-based or voice-assisted interaction. However, few DVI functions are presently controlled entirely through voice commands. At minimum, most current voice-based in-vehicle systems are multi-modal in nature, drawing upon at minimum a combination of auditory, vocal, cognitive, manipulative, and visual resources.

Research has been directed for some time at developing an understanding of and assessing the safety, usability, and demand related aspects of voice-interaction in the vehicle (see Barón & Green, 2006; Ei-Wen Lo & Green, 2013; Reimer, Mehler, Dobres, & Coughlin, 2013 for reviews). However, as detailed in Reimer et al. (2013), only limited publically available research was identified (seven studies covered in five reports) that addressed the evaluation of the demands associated with a driver engaging with production level, embedded in-vehicle voice command interfaces. Carter and Graham (2000) evaluated 32 participants' behaviors during a laboratory tracking task while completing a series of voice interactions with a Jaguar S-type voice interface. Harbluk, Burns, Lochner and Trbovich (2007) utilized the lane change task in two studies, each comprised of 16 participants, to assess drivers' behaviors with a 2005 Acura-TL voice system. In a sample of 25 manufacturer's employees who reported being regular SYNC users, Shutko, Mayer, Laansoo and Tijerina (2009) investigated drivers' behaviors with a 2008 model year SYNC system. In the first report of an on-road assessment of the demands associated with a voice interface, Chiang, Brooks and Weir (2005) reported on 10 drivers' (mostly engineers) behaviors with a 2004 Accord and 2005 Acura RL voice system. In a more recent report, Ownes, McLaughlin and Sudweeks (2010) report on the results of an on-road evaluation of 21 drivers interacting with a 2010 Mercury Mariner's voice system.

In response to the sparse literature on drivers' behaviors with production level voice systems and the National Highway Traffic Safety Administration's (NHTSA) pending development of Phase III Distraction Guidelines (auditory-vocal human-machine interfaces) (National Highway Traffic Safety Administration, 2013), the MIT AgeLab, with CSRC support, undertook a comprehensive assessment of a production-level voice command interface. This research aimed to add to the available body of information on the demands associated with the utilization of a production level voice interface in real-world highway driving conditions. In 60 participants, evenly distributed by gender across a younger (20-29 years) and relatively older (60-69 years) age group, data was collected on visual behavior, physiological arousal as a measure of cognitive demand, driving performance measures, and self-reported workload across a dozen tasks representing a variety of in-vehicle activities (Reimer, et al., 2013). These activities included preset radio tuning and a CAMP style manual radio tuning task (Angell et al., 2006) conducted both through traditional visual / manual controls and through the use of the voice command interface. Other voice command activities included full alpha-numeric destination address entry, route cancelation, song selection from a USB connected media storage device, a deliberate failure condition (song selection failure), contact dialing, and three levels of a cognitive workload benchmark task (n-back).

The results showed that cognitive processing demands / workload, as assessed by drivers' physiological responses (heart rate and skin conductance levels), were lower than initially anticipated by the lead investigators prior to initiation of the project Specifically, demands were all at or below those obtained during performance of the 1-back task, an activity that appears to place moderate cognitive demands on the driver. The results also showed that, except for the deliberate failure condition, self-reported workload for all voice-command involved tasks was nominally below that of the manual radio tuning task. As would be anticipated of a voice-command DVI, the visual demands associated with using the voice interface for the radio tuning were lower than those observed during classic manual tuning. However, glance analysis of other tasks, and particularly use of the voice interface to enter a full destination address into the navigation system, was markedly higher than anticipated prior to initiation of the study in terms of total off-road glance time. Specifically, the mean total off-road glance time during the address entry task was 32.8 seconds for the sample as a whole (25.9 seconds for younger adults and 41.7 for older adults). If current guidelines for visual demand associated with visual-manual interfaces were applied to this voice-command involved interaction, this level of visual demand would fail to meet both NHTSA's new 12-second maximum eyes off-road criteria (National Highway Traffic Safety Administration, 2013) as well as the Alliance of Automobile Manufacturers' (The Alliance) 20 second threshold (criterion 2.1.A) (Driver Focus-Telematics Working

Group, 2006). Voice-based dialing and song selection also showed relatively long total off road glance times, although not as marked as in the address entry task.

Finally, a number of older drivers were observed to engage in what could be characterized as Orienting Responses (ORs) during voice-command interactions. These are instances in which drivers spoke directly to the graphical user interface, oriented their bodies towards it, or acted in a way that suggested the voice system was perceived to be "in" the display screen situated in the center static.

Current Research Objectives & Approach

This report summarizes the results of a second CSRC-supported study that was originally conceived as focusing on evaluating the extent to which highly structured introductory training, or the lack thereof, in how to use a current production level voice-command system impacts a driver's willingness to try to use the system, ability to successfully use the system, and the level of workload and distraction associated with using the system. This evaluation was considered relevant to the study of voice-command systems given that most users avoid reading formal user's manuals or experience limited, if any, guided instruction on the operation of a voice system at time of purchase. In essence, the self-training condition was intended to assess many of the attributes of what a driver might encounter in a self-guided "driveway" training experience or when renting an unfamiliar vehicle. In addition, as the nature of the results of the first study became apparent during the design of the current work, determining the extent to which key findings of Study 1 (Reimer, et al., 2013) replicate, became an at least equally import objective.

The basic research design divides participants into two groups: a "structured training" group that received the same level of detailed training in the parking lot on how to use the voice-command system that was provided in Study 1, and a "self-trained" group. The "self-trained" group was given a list of the tasks that they would be asked to attempt during the drive and provided with the opportunity to explore the voice-command system on their own in the parking lot prior to the drive. Access to the user's manual was provided, but no other training support beyond experience with how to use the voice-command button as part of the procedure to calibrate the system to an individual participant's voice was given.

The primary assessment considered three task classes used in the first study: tradition manual operation of the radio, voice-command control of the radio, and voice-command control of the navigation system consisting of entry of a destination address and route cancellation. We deliberately reduced the number of DVI assessments explored in Study 1 so that the self-trained group was not unrealistically asked to explore too many functions on their own in a single experience. The multi-level MIT audio presentation – verbal response n-back task was again included as a cognitive demand, calibration reference.

A number of other characteristics of the study were adjusted to better align with current NHTSA DVI assessment guidelines, to reduce the overall time involvement of participants, and assess the impact of simply listening to audio stimuli. In particular:

- Our previous work examined a sample of younger (20-29) and older (60-69) drivers. The present study examines a sample of drivers equally distributed across four age groups: 18-24, 25-39, 40-54, and 55 and older. These age strata conform to NHTSA's recent recommendations for test samples (National Highway Traffic Safety Administration, 2013).
- The radio manual tuning task was adjusted to better conform to the reference task recommended by the Alliance of Automobile Manufacturers (Driver Focus-Telematics Working Group, 2006) and NHTSA (National Highway Traffic Safety Administration, 2013).
- The on-road task evaluation prompts for the route cancel task were altered from that used in Study 1 to remove explicit prompting on how to execute the cancelation command during the drive.
- The song selection task and contact dialing task were dropped from the protocol to reduce time and task training demands on participants.
- A blank-back form of the n-back cognitive demand reference task that asks participants to simply listen to a set of digits was added (i.e. the task has no memory or verbal response components).

Other significant design characteristic changes from Study 1 consisted of the following:

- Discontinued the use of the questions about the use of anti-anxiety and anti-depressant medications as part of study exclusion criteria - a recent review of people applying to be participants in studies at the AgeLab revealed that over 20% of females over the age of 50 who responded to on-line preliminary screening assessments had been placed on anti-anxiety or anti-depressant medications for various reasons by their physicians. Often these prescriptions appear to be for reasons other than for the treatment of clinical levels of anxiety or depression. It thus appeared that these questions were not serving the intended purpose of identifying possible health / safety concern outliners and were, instead, excluding individuals who are representative of a significant portion of the older population. We concluded that other screening items, such the question regarding whether an individual is under active treatment for a mental health condition, might more effectively fit the original safety screening goal.

- Discontinued the use of the Montreal Cognitive Assessment (MoCA) battery as a cognitive screener for participation for reasons detailed in the Study 1 technical report.
- Participants in Study 1 were given prompts that indicated the expectation that participants would engage in the various tasks during the driving portion of the study unless they explicitly felt it was unsafe to do so. Statements were phrased along the lines of "You will be asked to do two kinds of tasks". It is an open question if this type of prompting results in participants engaging in activities that they may otherwise prefer not to. In this study, where participants were being differentially introduced to / trained on the tasks, it is potentially interesting to assess the degree to which participants feel comfortable and/or inclined to engage in a voice-command system interaction as a function of how they have been introduced to and trained in the use of the system. We consequently modified slightly the prompting of participants and used phrases in the form of "while we would like you to consider doing each of the tasks¸ choosing not to do a particular task is totally acceptable…" to give participants more of a sense that it is ok to decline to engage in a particular task if they feel uncomfortable in doing so for any reason. Thus, declining to engage in a system interaction as a function of training category was considered a more overt variable in this study

Methods

Participants

Recruitment drew from the greater Boston area using online and newspaper advertisements and consisted of four age groups: 18-24, 25-39, 40-54, and 55 and older to comply with NHTSA guidelines on the testing of visual-manual distraction associated with DVI use. Participants were required to read and sign an institutional review board approved informed consent form, to present a valid driver's license and attest to having had their license for more than three years, to driving on average three or more times per week, and be in self-reported reasonably good health for their age. An experimenter verified that participants clearly understood and spoke English. Individuals were excluded if, on the basis of self-report, they had been involved in a police reported crash in the past year, had a major medical illness resulting in hospitalization in the past 6 months, had a diagnosis of Parkinson's, Alzheimer's, dementia, mild cognitive impairment, or other neurological problem, were being treated for a psychological or psychiatric disorder, had a history of heart failure, angioplasty, coronary artery bypass grafting, a pacemaker, stroke, transient ischemic attack, or diabetes. Medication exclusions consisted of the use in the past twelve months of anti-convulsants, immunosuppressive, cytotoxic, anti-depressant, anti-psychotic, anti-anxiety drugs, or medications to treat a major medical condition such as cancer. Also considered was the use in the past two days of any medications causing drowsiness. Potential participants were informed that the expected duration of the study was four to four and a half hours, including approximately two hours of on-road driving. Compensation was $90.

General Inclusion Criteria

- Age: 20 to 69 years
- A driver's license for more than 3 years
- Drive 3 or more times a week (on average)
- Comfortable speaking and reading English

General Exclusion Criteria (based on self-report):

- A driver in a police reported crash in the past year
- Failure to positively endorse the statement "Would you be comfortable driving a full size sedan" as part of the study.

- Failure to positively endorse the statement "Are you in reasonably good health for your age?" or if self-rating of health on in-lab screening questionnaire as "poor".
- Any major illness resulting in hospitalization in the past 6 months
- Diagnosis of Parkinson's, Alzheimer's disease, dementia, mild cognitive impairment (MCI), or any other neurological problems?
- Current treatment for a psychological or psychiatric disorder
- Report of having ever had heart failure, angioplasty or coronary artery bypass grafting (CABG), a pacemaker, stroke or transient ischemic attack, diagnosis of diabetes
- Use in the past 12 months of anti-convulsant, immunosuppressive, cytotoxic, anti-depressant, anti-psychotic, anti-anxiety medications
- Medication to treat a major medical illness (such as cancer) in the past 12 months
- Use of medication that made them drowsy in the past 2 days

Apparatus

The study was conducted in a 2010 Lincoln MKS with factory installed voice-command systems (Ford SYNC™ for voice control of the phone and media connected by USB and the "next-generation navigation system" with Sirius Travel Link). This was the same vehicle employed in Study 1 and was originally selected for study as the vehicle was already owned by MIT, instrumented for on-road research, and represented an example of a widely available production level voice interface when this project was initiated in 2011. The voice interface is engaged using a "push-to-talk" button on the right side of the steering wheel (see Figure 1). When the voice control interface is active, a display screen in the center stack typically supplies supporting visual information on system status and often provides information on prompts that the driver may use in dialog with the system (see Figure 2). A voice recognition training option is available in the system to optimize system capacity to recognize commands from an individual driver. This system training feature was utilized when a participant was introduced to the system to maximize the capacity of the system to correctly recognize commands from each participant.

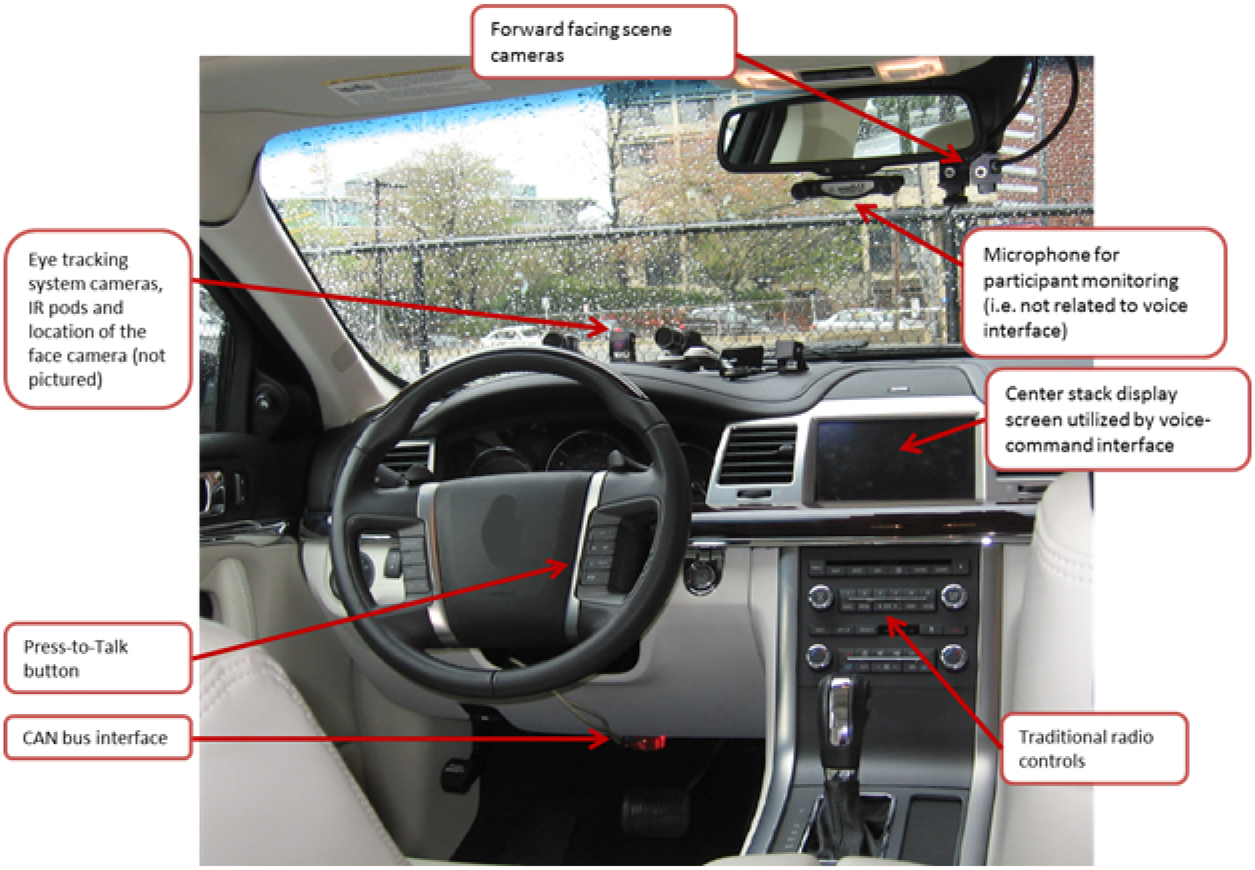


Figure 1: Interior of the test vehicle. Note the Push-to-Talk button on the right side of the steering wheel that is used to initiate interaction with the voice-command system and the center stack display screen (see image below).

Figure 2: The screen above appears on the display screen at the top of the center console when the Push-to-Talk button is pressed.

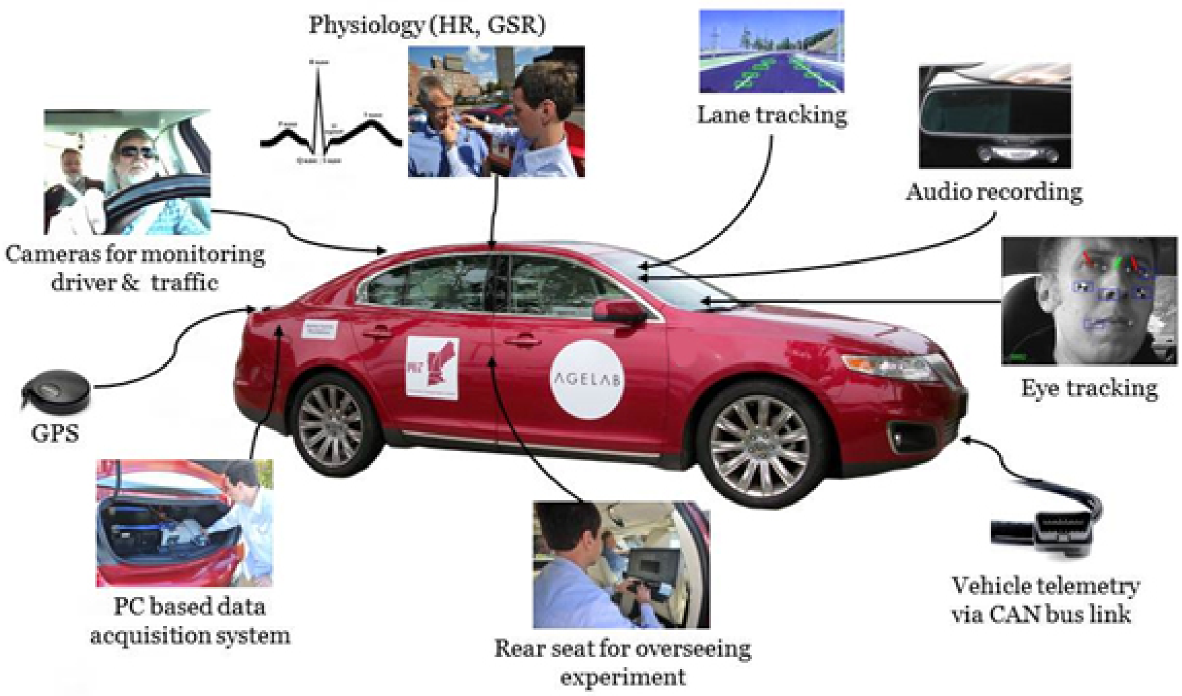


Figure 3: Experimental vehicle with key components noted. The identifying graphics shown on the side of the vehicle are removed during experimental sessions to avoid drawing attention to the vehicle and driver which might potentially impact normal traffic flow and interaction.

As graphically visualized in Figure 3, the vehicle was instrumented with a customized data acquisition system for time synchronized recording of vehicle information from the controller area network (CAN) bus, a MEDAC System/3 physiology monitoring unit, FaceLAB® 5.0 eye tracking, a microphone, an Iteris AutoVue® Lane Departure Warning System for assessing lane position, and GPS tracking. CAN bus and lane position data were captured at 10Hz, GPS data at 1Hz, physiological data at 250Hz to support EKG feature extraction for accurate heart beat interval detection, and eye tracking data was recorded at up to 60Hz. The camera configurations for capturing driver behavior and vehicle surroundings (see Table 1) was modified from Study 1 to include both color and black & white cameras, with the later optimized for coding eye glance behavior.

Table 1: Camera configurations and description.

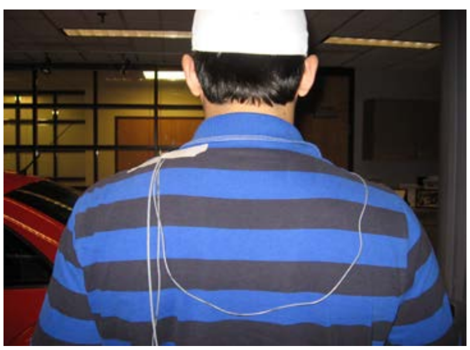
Camera Name / description	Frame Rate (fps)	Color	Image Size	Camera type / Lens
Forward View	30	N	640x480	Guppy Pro F125C/ Fujinon DF6HA-1B (6mm)
Forward View Wide Angle	15	Y	1024x240	Guppy Pro F125C/ Kowa LM4NCL (3.5mm)
Driver Face	15	Y	640x480	Guppy F033C/ Kowa LM6NCM (6mm)
Driver Face B&W	15	N	640x480	Guppy F033B/ Kowa LM8JC (8mm)
Over the Shoulder Dash	15	Y	640x480	Guppy F033C / Kowa LM6NCM (6mm)
Rear	15	Y	640x240	Guppy Pro F125C/ Kowa LM4NCL (3.5mm)

EKG recordings employed a modified lead II configuration; the negative lead was placed just under the right clavicle (collar bone), the ground lead just under the left clavicle, and the positive lead on the left side over the lower rib. The skin was cleaned with isopropyl alcohol and standard pre-gelled silver/silver chloride disposable electrodes (Vermed A10005, 7% chloride wet gel) were applied.

Skin conductance was measured utilizing a constant current configuration and non-polarizing, low impedance gold plated electrodes that allow electrodermal recording without the use of conductive gel. Sensors were placed on the underside of the outer segments of the middle fingers of the non-dominant hand and secured with medical grade paper tape. The thin surface design of the electrodermal sensors minimized interference with a natural grip of the steering wheel associated with the use of more traditional cup style electrodes. All wires were taped to participants for safety and positioned to allow for free movement (see Figure 4 for illustration of sensor placement).

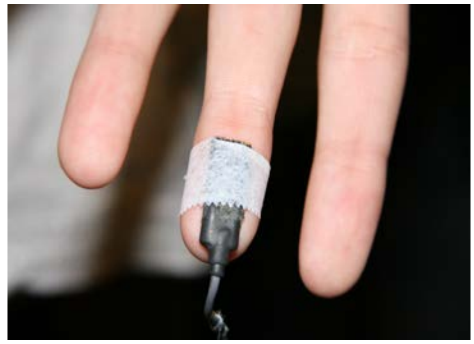
EKG Sensors (3 contacts) – Employing a modified lead II configuration, active leads were placed just under the right collar bone and over the bottom rib on the left side of the body create a vector across the heart. The sensor just under the left collar bone is the ground / reference. The skin is cleaned with alcohol and wiped dry before placing sensor.

EKG with blue dot on right side; orient cable up over right shoulder and gather together with left lead on the left shoulder as shown below:

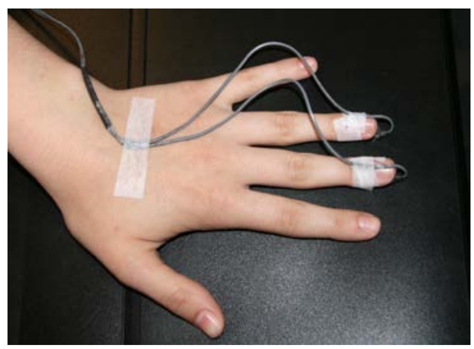


EKG placement over lower rib on the left side. Exact placement is not highly critical for this lead and it can be placed lower and somewhat farther back.

EDA Sensors for Skin Conductance (SCL) - Gold contacts were placed on the underside of the tip of the two middle fingers of the left hand. The inner edge of the gold contact is placed far enough forward so that the outer segment of the finger can bend normally around the steering wheel.



Lead wires are folded up and back over the top side of the fingers and held in place with medical paper tape.



To determine where paper tape should be placed on the back of the hand, the participant is asked to make a fist and draw their arm up toward their right shoulder as shown; the tape is then attached. Bending the fingers and elbow in this way corresponds to maximum pull that will occur on the lead wire.

Lead wire is taped at the 3 points shown and on top of shoulder.

Figure 4: EKG & electrodermal sensor attachment employed.

One of the outcomes of Study 1 was a conclusion that reliance on the FaceLAB® automated eye-tracking system under on-road driving conditions did not provide the level of reliability that we considered necessary for critical evaluation of eye glance behavior that involved significant glance off the forward roadway, particularly discrimination of glances low to the instrument cluster / steering wheel region as distinct from glances low to the center stack region. Thus, video recording of the driver's face was selected as the primary method for collecting data relative to eye glance behavior in the current study. Nonetheless, it was decided that automated eye-tracking data would collected during the study to allow for the possibility of further comparative evaluation of automated and manually-coded eye glance metrics at a future date.

As in Study 1, FaceLAB® calibration was performed, following the manufacturer's suggested procedures, as follows. Participants were instructed to sit in the driver's seat of the vehicle and look straight ahead. Two cameras mounted on the dashboard captured an image of the participant's face (this was reduced internally by the eye tracking system's internal algorithms to a representation of the participant's eye and facial features). From these data, FaceLAB® generated a model of the participant's face and eyes and tracks the changes in the positioning of these features in relation to a virtual "world model" environment that approximated the layout of the vehicle cabin. To verify that the eye tracker was properly calibrated, participants were asked to make a series of overt glances to objects of interest in the vehicle (i.e. the speedometer, rearview mirror, center stack touch screen, and finally, straight toward the front windshield). If FaceLAB®'s estimates of these glance targets were not accurate, the experimenter adjusted the positioning of world model objects was until the system produced an "observed" accurate estimate of the participants gaze positioning. The system was re-calibrated daily through the manufacture specified producers, "picture of a checker board calibration at various angles" to ensure that the system internal representation of the camera positions remained accurate.

An experimenter was seated in the rear of the vehicle and was responsible for providing driving directions, ensuring safe vehicle operation, that participants understood and followed instructions, recording telemetry was working properly and that the experiment proceeded according to a predefined script. The data acquisition system supported playing recorded audio and the experimenter used a set of F-key presses at predefined points to trigger steps in the experiment. This ensured that primary instructions and tasks were presented in a consistent manner.

DVI & Surrogate Cognitive Workload Tasks

There were three in-vehicle task areas assessed: manual control of the radio, voice command control of the radio, and navigation system destination entry and route cancelation. In addition, four levels of a cognitive demand calibration task (n-back) were considered. Illustrations of selected tasks appear at the following links:

Manual Radio Tuning - http://youtu.be/7Gss2bicmNE
Voice Radio Tuning - http://youtu.be/95p4ML-oZu0
Voice Navigation Entry - http://youtu.be/qDklZX5FUR0
N-Back Calibration Tasks - http://youtu.be/qu4OTH1REHs

Radio Tasks

As detailed in Reimer et al. (2013), basic radio interaction considered in Study 1 was modeled on protocols developed as part of the Crash Avoidance Metrics Partnership (CAMP) Driver Workload Metrics Project (Angell et al., 2006). Following the approach taken by CAMP, the “hard” version of the manual radio task consisted of turning the radio on, selecting a radio band, and then tuning to a specified station by rotating a manual tuning knob. The radio turning reference task adopted by The Alliance (2006) and recently endorsed by NHTSA, is potentially slightly easier than the CAMP “hard” task in that it assumes that the radio is already on. The radio “hard” task was revised for Study 2 to specifically correspond to the characteristics NHTSA specifies for the radio tuning reference task for assessing visual-manual DVIs. (A detailed listing and discussion of the specific differences between the form of the tasks employed is provided in *Appendix B: Radio Task Changes from Study 1 to 2*).

Task Execution Notation:

[button]	a button press; i.e., press voice command button [voice] on steering wheel, hard button on console, or soft button on touch screen
(tune)	rotate manual tuning knob
“yes”	say a voice command

Radio Easy task: Radio is on, change radio station to Preset-X. For the manual task, the Preset buttons are classic style, numbered hard physical buttons as shown in Figure 5.

Manual - 1 step	**[Preset-1]**
Voice - 3 steps	**[voice] → “Preset-1” → “yes”**

Radio Hard task: Extant NHTSA guidance specifies that for visual-manual distraction assessment using the radio reference task, the radio is to be on at the start of the task. Further, if the radio controls are part of an integrated vehicle display, the integrated display should be set so that the radio controls are not active and the participant is to perform the action(s) necessary to make the radio controls active. The participant is then to switch between AM and FM bands, and finally manually tune the radio to a specified station using available rotary tuning or seek buttons. In the test vehicle used in the current study, the [RADIO] button is a physical (hard) button located directly below the preset-1 button (see Figure 5). Pressing the [RADIO] button results in the display of band options: AM, FM1, FM2, Sat1, Sat2, and Sat3 on the center console display screen (see Figure 1). A desired band may then be selected by placing a finger on one of the touch screen (soft) buttons.

Manual – min. 3 steps	**[RADIO] → [AM] → (tune)**
Voice – 3 to 4 steps	**[voice] → "100.7" → "yes"**
	[voice] → "Radio" → "100.7" → "yes"

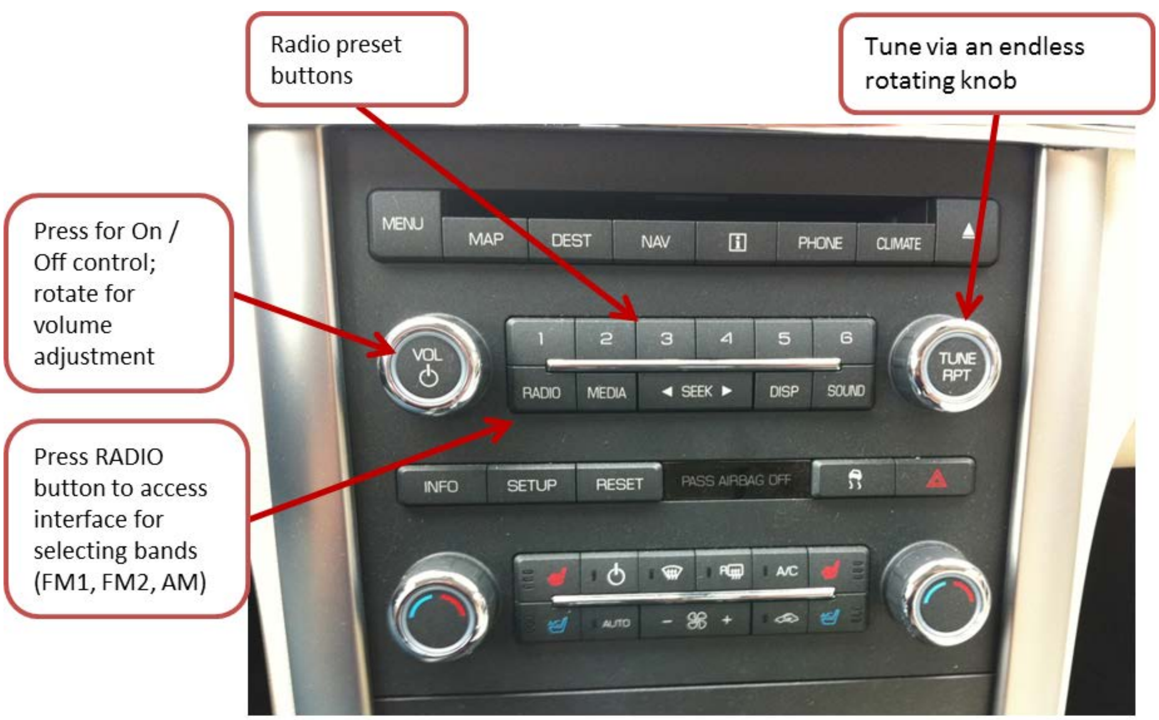


Figure 5: "Traditional" visual-manual radio interface in 2010 model year Lincoln MKS test vehicle. Switching between FM1, FM2, and AM radio bands was carried out using touch screen buttons on the main display screen located directly above the center console shown here.

Use of soft buttons for band selection was chosen in the guided training condition over an available option that allows users to toggle between bands by multiple presses of the [RADIO] button. This selection was made for two reasons. First, the soft buttons were highly salient once the [RADIO] button was pressed and likely to be the default option chosen by many participants even if instructed otherwise. Second, since no subscription to the satellite radio system was purchased, when toggling past those bands a message was presented on the touch screen noting the lack of subscription. This message needed to be closed by a press of the touch screen to move past each of the three Sat station bands. Self-trained participants were given no specific guidance on the method to use to switch bands.

The radio stations selected in the current study varied somewhat from those used in Study 1 to support switching between the AM and FM (FM1 & FM2) bands and then from FM to AM. (Study 1 used the sequence AM to FM2 and later switching to FM1 The steps in both studies were functionally equivalent but the change to AM to FM, and then FM to AM was made to correspond literally to specifications provided by NHTSA. (See *Appendix B: Radio Task Changes from Study 1 to 2*for additional detail.)

Manual Radio Tasks

EASY 1: Your task is to change the radio to **preset-1**

EASY 2: Your task is to change the radio to **preset-6**

HARD 1: Your task is to switch to **AM**, and tune to **1470**.

HARD 2: Your task is to switch to **FM**, and tune to **100.7**.

Voice Radio Tasks

EASY 1: Your task is to change the radio to **preset-1**

EASY 2: Your task is to change the radio to **preset-6**

HARD 1: Your task is to request **AM 1470**.

HARD 2: Your task is to request **FM 100.7**.

Navigation Tasks

Voice-command interaction with the navigation system consisted of two subtasks, entry of a street address and cancelation of the route request. Assuming there were no overt errors in interaction with the system, address entry required between 12 to 16 discrete steps. The number of steps appeared to vary depending on the confidence level of the system on the recognition of a voice entry and the extent to which a given street or town entry had variant options that the system required the user to select from. The same addresses were presented in Study 1 and 2.

In Study 1, the Navigation Cancel task was deliberately structured to be as low demanding as possible. Consequently, the task request instructions given to the participant prompted the specific voice-command that was to be used to cancel the route request (i.e. "*Your task is to cancel the route using the command '**Navigation Cancel Route**'.*"). Given that one of the design goals for Study 2 was to compare the behavior of participants who were given guided training in how to complete tasks with that of participants who attempted to learn how to complete tasks on their own, the evaluation prompt during the driving phase of the study was changed to drop the specific description of the command to be used to cancel the route. Specifically, the form of the task was changed to "*Your task is to cancel the route you entered*". As detailed in the results, this change resulted in most participants finding this task to be appreciably more challenging to complete than was the case in Study 1. (See Appendix C for additional detail on differences between Study 1 and 2 as well as full text of the recorded audio prompts employed.)

Voice Navigation Tasks

HARD 1: Your task is to enter the destination address: **177 Massachusetts Avenue, Cambridge, Massachusetts**.

EASY 1: Your task is to cancel the route you entered.

HARD 2: Your task is to enter the destination address: **293 Beacon Street, Boston, Massachusetts.**

EASY 2: Your task is to cancel the route you entered.

Task Execution:

Address Entry (Nav "Hard") - 12 to 16 steps – variable depending on speech recognition and whether listing(s) of city or street name selections to choose from was presented by the system.

Driver	**System**
[Presses the voice button]	*Please say a command.*
"Destination Street Address"	*Destination Street Address, is that correct?*
"Yes"	*Destination Street Address, please say the city name.*
"Cambridge"	*Cambridge, is that correct?*
	[If system is unsure of the city name entry, a listing of options is presented on the display screen and driver is asked to say a line number.]
"Yes"	*Cambridge, Please say the street name.*
"Massachusetts Avenue"	*Massachusetts Avenue, is that correct?*
	[If system is unsure of the street name entry, a listing of options is presented on the display screen and driver is asked to say a line number.]
"Yes"	*Massachusetts Avenue, please say a house number.*
"One Seven Seven"	*The number you said was 1-7-7, is that correct?*
"Yes"	*Yes. Press the voice button to say set as destination or back.*
[Presses the voice button]	
"Set as Destination"	*Set as destination, is that correct?*
"Yes"	*Set as destination.*

Cancel Route (Nav "Easy") – min. 3 steps:

Driver	**System**
[Presses the voice button]	*Please say a command.*
"Navigation Cancel Route"	*Set as destination, is that correct?*
"Yes"	*Set as destination.*

N-Back Surrogate Tasks

In 2006, the MIT AgeLab began a project to assess the feasibility of biometric-based state detection under driving conditions that considered a wide range of possible physiological measures. To carry out this evaluation, a method of reliably inducing multiple levels of arousal or demand was required. A number of different methods were considered and eventually a variation of a cognitive task widely used in neuropsychological and medical research was adapted for use in the project. The resulting n-back task variation developed in the AgeLab requires participants to hold single digit numbers in memory and to repeat them back verbally either immediately (0-back), after another number has been presented (1-back), or after two additional numbers have been presented (2-back). As shown in the example below (Table 2), the numbers are presented as a random ordering of the digits 0-9 with a typical spacing of 2.25 seconds between numbers. Single 10 item stimulus sets were employed in this study, resulting in task periods of approximately 30 seconds in duration. A new form of the n-back was added to this study called the “blank-back”. For this condition, participants were instructed to simply listen to each number as it was read. As noted below, the intent of adding this condition was to assess the impact on participants of simply attending auditorily without the need to verbally respond or hold in memory.

Table 2: Example of an N-back task set.

Stimulus	6 9 1 7 0 8 4 3 5 2
0-back Response	6 9 1 7 0 8 4 3 5 2
1-back Response	. 6 9 1 7 0 8 4 3 5
2-back Response	. . 6 9 1 7 0 8 4 3
Blank-back	. . . .

Note: During the “blank-back” the participant is asked to attend to the number but not say anything.

As can be seen from the table above, for the 0-back task the participant simply has to repeat each number as it is presented. In the 1-back task, the participant is to hold a number in memory, wait for the next number to be presented and then enter it into memory, and then verbalize the previous number while continuing to hold the most recent number in memory. The 2-back extends upon the 1-back by requiring that the participant hold the most recent two numbers in memory. The vocal demands of this task are relatively consistent, with the 1-back requiring one less vocalization than the 0-back and the 2-back requiring one less vocalization than the 1-back. Consequently, the task largely represents a manipulation of the level of demand on working memory.

Extensive research has been undertaken on the use of a delayed digit recall task (n-back) as a method for inducing graded levels of cognitive demand during simulation and actual on-road driving (Mehler, et al., 2012; Mehler, et al., 2009; Reimer, 2009; Reimer & Mehler, 2011; Reimer, Mehler, Wang, & Coughlin, 2012; Son et al., 2011).

Blank-back Task – In addition to the established 0- through 2-back conditions, a “blank-back” condition was added in the present study where participants simply listen to a series of single digit numbers (like the other n-back levels) but in which participants do not have to hold in the number in memory or repeat it out loud. The intent of adding this condition was to assess the relative level of physiological arousal that is associated with the simple auditory presentation of stimuli without any memory demand component. This should add to our understanding of how basic listening to auditory stimuli impacts the driver (driving performance, physiological arousal, eye behavior) relative to the more demanding cognitive aspects present in the classic n-back listening/memory/vocal-response task.

Floating Reference Period – In addition to adding the blank-back condition, the present study included a floating “single task” reference period randomly interspaced with the four levels of the n-back task in which the participant was prompted to “just continue driving”. This condition provides a reference period of equal duration of the n-back periods and allows for an estimate of the impact of the participant being presented with an auditory prompt. Even though the prompt only requests the participant to continue driving, hearing a prompt could potentially have some impact on various aspects of a driver’s behavior.

Exploratory Pacing Task – Previous work by our group on the n-back task has always utilized a fixed-pacing approach to stimulus presentation. A variant of the typical 1-back task was included in the data collection with the intent of exploring the relative impact of the standard fixed-paced item presentation with user-controlled, self-paced interaction. As detailed below, this task was always presented on the return trip back to MIT following the completion of the primary study tasks. This portion of the protocol was intended to collect additional basic research level data on the relative demand of self-paced vs. system paced voice interaction. Analysis of this data was not planned as part of the formal project deliverables and is not presented as part of this report. (Informal review of this portion of the dataset suggests that the methodology employed for this task may not have been optimal and we will most likely revisit this question in future work.)

Procedure

As in Study 1, detailed protocol checklists provided instructions for the experimenter including language guidance for all key interactions with the participant. Pre-recorded audio instructions were used at key points during training to ensure that important concepts were presented in a consistent manner across participants and all task specific prompts during the on-road evaluation were also presented using pre-recorded audio to ensure consistency.

Outline of Study Protocol

In-Lab Start Phase

- When Participant Arrives - Consent Forms / Payment Form / Emergency Contact Form
- Review of Eligibility (Interview)
- Pre-Experimental Questionnaire (Parts I & II)
- Visual short-term memory (VSTM) assessment (BrainBaseline)
- N-Back Training
- Workload Scale Rating Explanation

Bathroom Break

- Physiological Sensor Attachment

Move to Vehicle

- Set Participant Up in Vehicle / Eye Tracking Calibration / N-Back Practice
- Participants follow either a “training” or “no training / self-exploration” path – manual radio, voice radio, navigation system practice periods
- Start On-Road Run to I95 North (habituation)
- I495 South– Random ordering of 4 primary tasks

Task 1 – One of the following task categories: **voice-radio**, manual-radio, voice-nav, or fixed-paced n-back (*random ordering without replacement, i.e. each task category was only presented once during the drive.)*

Task 2 – One of the following task categories: voice-radio, **manual-radio**, voice-nav, or fixed-paced n-back.

- Turn around – optional bathroom break if needed – I495 North

Task 3 - One of the following task categories: –voice-radio, manual-radio, **voice-nav**, or fixed-paced n-back.

Task 4 – One of the following task categories: voice-radio, manual-radio, voice-nav, or **fixed-paced n-back**.

- Return to MIT on 93 S

Task 5 – 2 trials each of fixed-paced 1-back and self-paced 1-back (4 trials total) (half Ss get fixed-paced first, half self-paced first). (*This was an exploratory protocol and not directly related to the primary study; analysis and results are not included in this report.)*

Back at MIT

- Workload Rating Scale for all tasks in-vehicle
- Detailed Task Rating, Post-Experimental Q, and Supplementary Health Q

Following informed consent, a review of eligibility criteria, and completion of a pre-experimental questionnaire, participants were trained to minimal competency criteria on the n-back task as in Mehler, Reimer and Coughlin (2012) and then given an explanation of how to complete the workload rating scale (see Appendix A). A bathroom break was offered, physiological sensors attached, and participants were then escorted to and given an orientation to the research vehicle. The participant was instructed to adjust seat and mirrors and asked to back up the vehicle a few feet before picture were taken for calibration of the eye tracking system. Additional 10-item sets of each of the levels of the n-back task were practiced in the stationary vehicle as the RA configured the eye tracker (see procedures above). An introduction to the voice command system was provided that included going through the individual voice calibration option.

As detailed previously, three in-vehicle task areas were assessed: manual control of the radio, voice command control of the radio, and navigation system interaction consisting of destination entry and route cancelation. In addition, four levels of a cognitive demand calibration task (n-back) were considered. Each task type was presented twice. The surrogate n-back task was intended specifically as a reference cognitive task for calibration / comparison scaling against the primary system tasks. The four task types (manual radio control, voice command radio control, voice command navigation entry, and n-back were presented in random order across the sample. A general outline of task workflow and counterbalancing is shown in Figure 6where Task 1 could be any one of the four task types, etc.

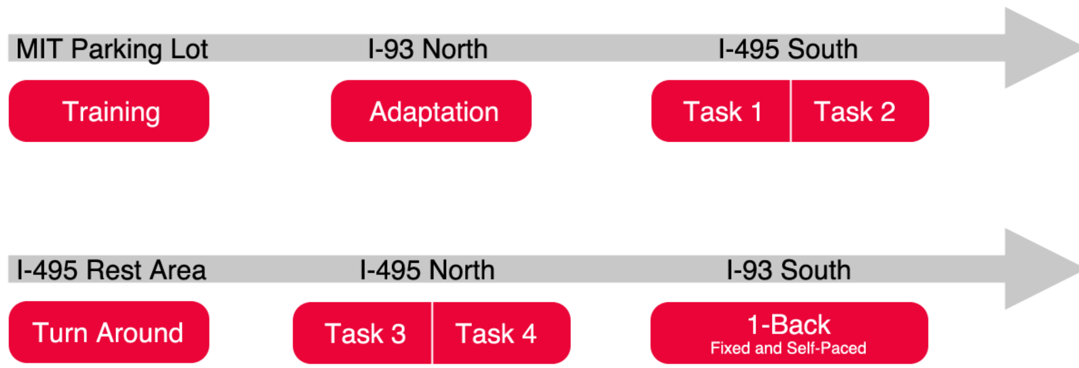


Figure 6: 'Task 1', 'Task 2', etc. represent the four primary task categories employed in this experiment: manual radio control, voice command radio control, voice command navigation entry, and n-back. Ordering of the task categories was randomized across the sample.

On-Road Assessment

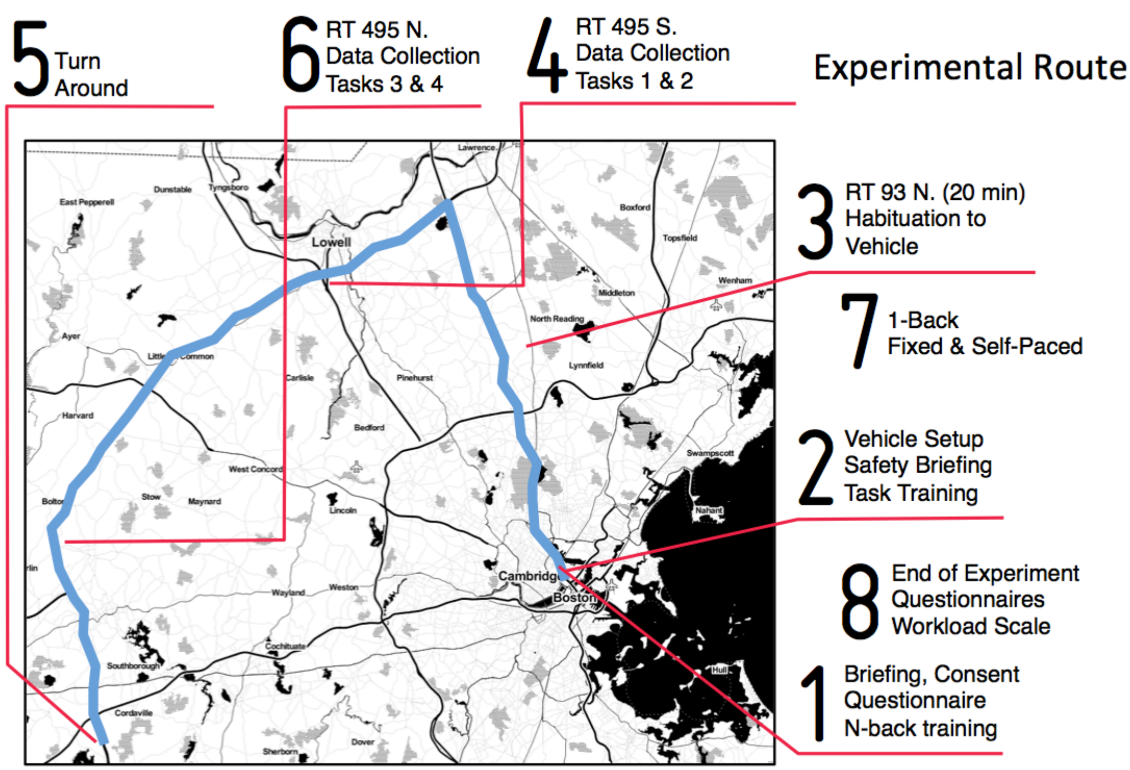


Figure 7: Experimental route with key protocol periods.

The driving portion of the study was conducted on roadways in the greater Boston area (see Figure 7) and divided into four segments. The first consisted of a period of approximately 10 minutes of urban driving to reach interstate highway I-93 and continued north on I-93 for an additional 20 minutes or so to the I-495 intersection (step 3 in the figure). This allowed a total adaptation period of approximately 30 minutes of driving prior to the assessment portion of the study. The second segment consisted of driving south on I-495 until the participant was able to complete the first two task periods (step 4 in the figure) and then reach a convenient exit point to reverse direction; the drive averaging approximately 30 minutes. The third was from the turn-around point back north on I-495 to I-93 during which the remaining two primary task periods were completed (step 6). The fourth segment consisted of the return on I-93 south where the fixed vs. self-paced 1-back tasks were presented (step 7).

The sub-task ordering for the radio and navigation task areas always followed the same presentation order specified earlier in the Methods. The ordering of the four levels of the n-back task was randomly distributed across the sample. For the exploratory fixed vs. self-paced 1-back task, presentation order of the fixed and the self-paced trials was counterbalanced across the sample. Self-report workload ratings were obtained at the completion of the drive in the MIT parking lot prior to the participant leaving the vehicle.

Measurements

Self-Report

Questionnaires

A pre-experimental questionnaire was employed to collect demographic information, driving history, technology experience, and other information. A post-experimental questionnaire was employed to gather information on participants' experiences using both structured and open-ended questions. (Analysis of these questionnaires is not included in the current technical report on primary findings. It is anticipated that exploration of this supplementary dataset will be integrated into future work considering drivers' experiences and impressions of working with voice-command interfaces.)

Self-Reported Workload Ratings

Subjective workload ratings were obtained using a single global rating per task on a scale consisting of 21 equally spaced dots oriented horizontally along a 10cm line with the numbers 0 through 10 equally spaced below the dots and end points labeled "Low" and "High" on the left and right respectively (see *Appendix A: Self-Reported Workload Materials* for details on the instrument, item wording, and reproduction of the scale). All of the scales were presented on a single sheet of 8.5x17 inch (legal size) paper so that participants were able to rate each task relative to tasks that they had already rated. Participants were told that workload may involve mental effort, the amount of attention required, physical effort, time pressure, distraction or frustration associated with trying to do the task while continuing to drive safely, and that workload is best assessed by the person doing the task. They were instructed to circle a point along each scale that best corresponds to how much workload they felt was involved in trying to do each task.

Physiological Metrics

Heart Rate

Heart beats were detected through identification of R-wave peaks in the EKG signal. Processed records were reviewed by trained RAs to identify and resolve any detection issues.

Skin Conductance Level (SCL)

High frequency noise in the skin conductance level (SCL) signal was removed through a wavelet transform (see Reimer & Mehler, 2011). Gross low frequency movement artifact was identified by manual inspection and removed.

Driving Performance Measurements

Measures of driving speed, acceleration data, and steering wheel metrics were recorded directly from the vehicle CAN bus.

Data on lane departure events was collected on an exploratory basis using an after-market Iteris AutoVue 3G lane departure warning system. Lane departure events were found to be extremely rare in Study 1 and appeared to be more overtly associated with task duration than the specific nature of the task. Due to the resource cost of manually reviewing video of each identified event and coding whether the event represented an intentional or unintentional lane departure, this variable was not included in the current analysis. The source data is being maintained to allow for potential future assessment of this variable.

Mean and Standard Deviation of Velocity

Two driving performance metrics included in this report are the mean and standard deviation of forward vehicle velocity, both measured in m/s. Input data for these metrics was obtained from the vehicle CAN bus.

Acceleration Events

CAN bus data of longitudinal and lateral acceleration was used to calculate independent acceleration events, as proposed in Reimer et al. (2012). This measure examines unidirectional acceleration, computed from individual lateral and longitudinal measures using the Pythagorean Theorem. Classification of independent acceleration events is parameterized with thresholds for both temporal separation and acceleration magnitude. For this report, an acceleration threshold of 0.1g (0.98m/s^2) and a temporal separation of 2 seconds between independent events were applied.

The count of acceleration events was normalized by each participant's trial duration, yielding the acceleration event rate, expressed in units per minute.

Steering Wheel Metrics

Steering wheel reversals were classified as proposed in the final report of the European Union AIDE project (deliverable D2.2.5, section 7.12) (Östlund et al., 2005) . This metric captures the number of steering wheel inputs exceeding an angular reversal gap of either 3° for major or 0.1° for minor reversal events. The rate of steering wheel reversals per minute was obtained by dividing the raw reversal rate by the task trial duration.

Additionally, the standard deviation of steering wheel angles, reported in angular degrees, was calculated based upon raw steering wheel angle information.

Glance Coding & Metrics

In-vehicle video for all participants and task periods of interest were coded to summarize participant glance behavior. Eye glance behaviors were coded manually with the assistance of software specifically developed by the AgeLab for this purpose. Two independent coders manually assessed video of each task. A third coder resolved any discrepant glance codes. Detailed procedures for glance coding and mediation are given in the technical report for Study 1 (Reimer et al., 2013). Specific metrics considered for this report are detailed below.

For purposes of the current report, primary glance metrics are quantified in terms of glances identified as eyes-off-the-forward-roadway (TEORT) following current guidelines specified by NHTSA (2013) as regard total eyes-off-road time (TEORT) determination. A detailed side-by-side comparison of the implications of using TEORT vs. the more traditional "glance-to-device" (GTD) method of quantification was presented in the technical report for Study 1 (Reimer et al., 2013). Selected GTD metrics and comparisons of the TEORT and GTD approaches are also provided in this report.

Mean Single Glance Duration
This metric considers the mean single glance duration away from the forward road scene for a specified period per participant.

Long Duration Single Glances
This metric considers the percentage of single glances per participant of defined types during a specified period that have durations greater than 2.0 seconds. Both NHTSA's TEORT and the GTD metric are considered.

Total Off-Road Glance Time
This considers the sum of the durations of each individual participant's specified eye glances for a specified period. Both NHTSA's TEORT and the GTD metric are considered.

Number of Off-Road Glances
This metric considers the number of glances of each individual participant away from the forward road scene for a specified period.

Orienting Response
This assessment of participants' behavior toward the center stack graphic display was developed during the post-data collection review of participant behavior during Study 1. The rating attempts to characterize a behavior in which the participant appears to engage directly with the in-vehicle display as if the voice-command interface were located within it. For example, the participant might begin speaking toward the display's location, lean towards it, change his posture, turn his body, or otherwise behave in a manner that suggests he has begun to prioritize interaction with the in-vehicle display. The detailed coding guide for orienting response behavior is given in the technical report for Study 1 (Reimer et al., 2013). Full coding of this behavior pattern was outside the scope of the original project plan for Study 2. However, an exploratory evaluation was carried out as detailed in the Results section.

Data Reduction & Analysis

Baseline reference values were computed for selected metrics as average values obtained across seven-two minute long single task driving periods. Each of these periods occurred immediately prior to the seven different task periods. In the case of driving performance data, these baseline periods are presented along with the metrics describing performance during tasks. In the case of physiological measures, task periods are presented as change scores from the baseline as well as in absolute heart rate and skin conductance values. Each task type was presented twice during the drive.

Statistical Analysis

Statistical tests for main effects were done in two parts. First, a full repeated-measures ANOVA model with factors of task period, age, gender, and training condition was performed. Since these tests showed the training did not significantly impact any measures of interest (see Results), a second ANOVA model was tested that disregards training condition and considers the combined groups as a single sample under study.

Statistical analyses were performed in R (R Core Team, 2013). Owing to non-normality of the sample data and /or the use of ratio data (percentages) for several dependent measures, in most cases non-parametric statistics such as the Wilcoxon signed rank test and the Friedman test were used (similar to the t-test and repeated-measures ANOVA, respectively). For selected analyses, repeated-measures ANOVA by ranks are presented.

Results

Sample Statistics & Screening Results

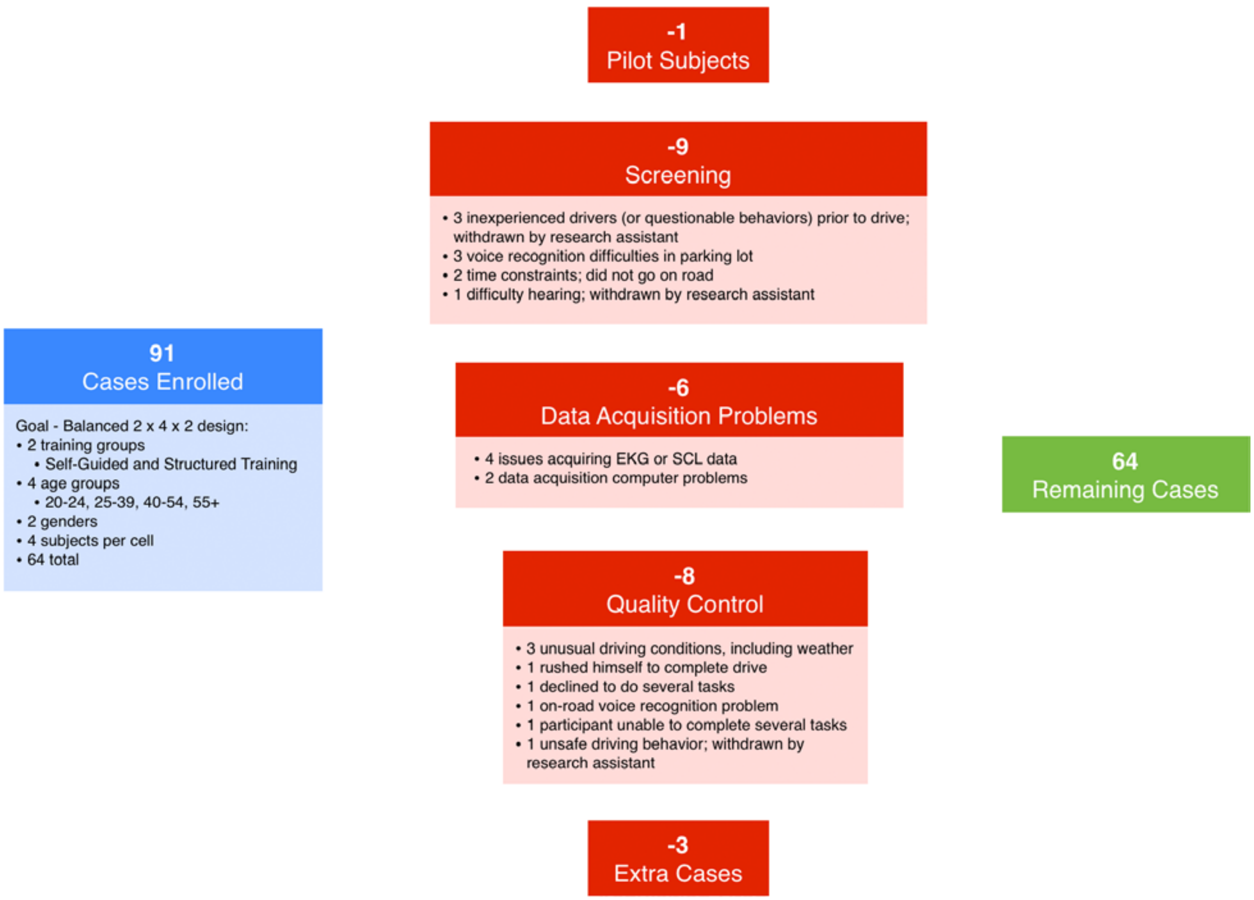


Figure 8: Break-down of participants recruited and flow through to final analysis sample.

The research plan called for recruiting a sample sufficient for obtaining a final analysis sample of 64 participant cases equally distributed by gender, across the four NHTSA age groups, and the two training conditions. As detailed in the figure above, 91 participants were initially enrolled in the study. Nine participants were dropped prior to going on-road for the reasons indicated in the figure. As in Study 1, overall voice-recognition capability of the system was relatively high, with only 3 participants being dropped due to basic voice recognition issues during the training phase in the parking lot. Six cases were dropped due to issues with physiological signal quality or computer related issues in data collection. Eight cases were excluded due to a variety of quality control issues such as heavy traffic, weather conditions, and inappropriate participant driving behavior. One participant declined to engage in several tasks during actual driving, feeling that they could not complete them safely. One participant had difficulty completing most of the tasks and was excluded. Three otherwise usable cases were classified as extra cases once the full 64 cell design was filled.

Summary statistics by age and gender are shown in Table 3. Age was closely balanced by gender across each of the age groups. Note that the demographics have not been broken down by training group as the type of training did not have an overall impact on the primary dependent measures considered in this report; see next section for additional detail.

Table 3: Demographic statistics. Each cell represents the mean (SD) [range] for 8 participants.

Age	Female	Male
20-24	22.12 (1.1) [20.0 - 24.0]	22.00 (1.3) [20.0 - 23.0]
25-39	33.00 (5.9) [25.0 - 41.0]	29.38 (5.3) [20.0 - 36.0]
40-54	46.75 (3.0) [41.0 - 50.0]	46.62 (3.7) [42.0 - 54.0]
55+	58.25 (1.8) [56.0 - 61.0]	59.50 (4.2) [55.0 - 69.0]

Summary Findings Comparing Training Conditions

The first formal question to be addressed in Study 2 (the current study) was whether the training method provided to participants (structured and detailed instruction vs. self-guided) significantly impacted any of the primary assessment metrics. Side-by-side task plots are provided in subsequent sections for each dependent measure. The comparison of results by training condition for task completion time is shown in Figure 9 below as an example of this time of plot. As can be observed for task completion time, the response patterning across tasks for both training conditions are remarkably similar for most of the outcome measures.

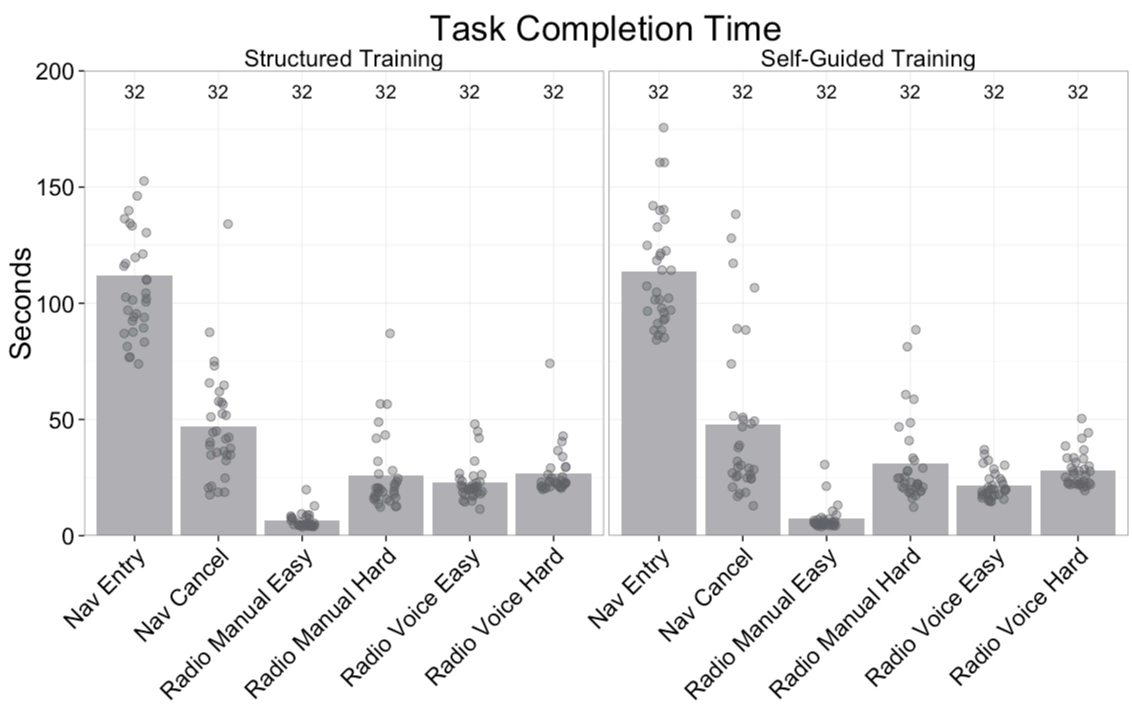


Figure 9: An example comparison plot showing task completion time by task and training condition.

A series of repeated-measures ANOVAs were conducted with training condition as a between subject factor and the DVI tasks (navigation address entry, navigation route cancel, "easy" manual radio tuning task, "hard" manual radio tuning task, "easy" voice radio tuning task, and the "hard" voice radio tuning task) as within subject factors. In contrast with initial expectations, structured training in how to carry out the tasks vs. self-guided training had no overall main effect across the tasks. As shown in Table 4 below, there was no significant main effect of task condition for any of the primary measures.

Table 4: Statistical Significance for Main Effects of Training Condition on Each of the Primary Outcome Measures

Measure	F statistic	p value
Self-Reported Workload	2.66	p = .109
Task Completion Time	1.31	p = .257
Heart Rate	0.00	p = .976
SCL	1.12	p = .295
Mean Velocity	0.02	p = .903
SD Velocity	0.03	p = .855
Acceleration Events (=>.1g)	0.10	p = .749
SD Steering Wheel Angle	0.02	p = .902
Minor SWR	0.44	p = .510
Major SWR	0.02	p = .902
TEORT Mean Glance Duration	1.38	p = .245
TEORT % Glances > 2s	0.59	p = .445
TEORT Total Glance Time	0.54	p = .466
TEORT Number of Glances	1.55	p = .219
GTD Mean Glance Duration	0.04	p = .839
GTD % Glances > 2s	0.33	p = .567
GTD Total Glance Time	1.74	p = .194
GTD Number of Glances	2.86	p = .097

TEORT = Total-Eyes-Off-Road-Time (NHTSA metric); GTD = Glance-to-Device; *statistically significant

The lack of a main effect of training condition across all of the dependent measures has two implications for the dataset. First, the general high level of consistency in the patterning of results increases confidence that the underlying patterns being observed represent relatively consistent phenomena. Second, this finding provides a rational for combining the two sub-samples into a larger single population of 64 participants for subsequent assessments.

Summary Findings Comparing Manual & Voice-Based Radio "Hard" Tuning

A key question in this project is whether a voice-command interface offers any objectively measureable advantage over a traditional visual-manual interface for accomplishing comparable tasks. In particular, voice interfaces are often presumed to offer a form of interaction that can reduce the visual demand associated with a traditional visual-manual DVI. Table 5 below summarizes the results of statistical tests that compare manual (m) engagement with the radio hard tuning task and the voice-command (v) based method of engaging with the same task on various glance based measures. The "+ Voice" column indicates measures for which the voice-command method might reasonably be interpreted as being less visually demanding or having a smaller impact than when engaging in the task using the manual interface. Descriptive statistics for each variable are provided in the sections on individual measures.

Table 5: Glance Metrics Comparisons for Manual vs. Voice-Based Radio Hard Tuning1 (Wilcoxon tests2) across Studies 1 (Shaded) & 2 (Unshaded).

Measure	+ Voice	- Voice	V statistic[2]	p value
TEORT Mean Glance Duration	m > v		1347	p **< .001***
TEORT Mean Glance Duration	m > v		2072	p **< .001***
TEORT % Glances > 2s	m > v		426	p **< .001***
TEORT % Glances > 2s	m > v		543	p **< .001***
TEORT Total Glance Time	m > v		1175	p **< .001***
TEORT Total Glance Time	m > v		1982	p **< .001***
TEORT Number of Glances			915	p = .079
TEORT Number of Glances	m > v		1920	p **< .001***
GTD Mean Glance Duration	m > v		1392	p **< .001***
GTD Mean Glance Duration	m > v		2069	p **< .001***
GTD % Glances > 2s	m > v		398	p **< .001***
GTD % Glances > 2s	m > v		542	p **< .001***
GTD Total Glance Time	m > v		1349	p **< .001***
GTD Total Glance Time	m > v		2024	p **< .001***
GTD Number of Glances	m > v		1283	p **< .001***
GTD Number of Glances	m > v		2015	p **< .001***

TEORT = Total-Eyes-Off-Road-Time (NHTSA metric); GTD = Glance-to-Device; *statistically significant

[1]As detailed in the text, the radio hard tuning tasks differed somewhat between Study 1 and Study 2, with the former following the CAMP format and the latter following the current NHTSA guidelines.

[2]The distribution of the V statistic for the Wilcoxon test is such that smaller and larger values are associated with the two ends of the probability distribution (and hence statistical significance).

As will be considered in more detail in the results sections that follow and in the discussion, it is evident from even a quick glance at Table 5 that the pattern of findings across the glance metrics and both studies is highly consistent. All of the metrics appear to support the position that utilizing the voice interface resulted in the driver being more oriented toward the forward roadway during the radio “hard” tuning task.

Table 6: Other Comparisons for Manual vs. Voice-Based Radio Hard Tuning[1] (Wilcoxon tests[2]) across Studies 1 (Shaded) & 2 (Unshaded)

Measure	+ Voice	- Voice	V statistic[2]	p value[3]
Self-Reported Workload	m > v		619	p = **.036***
Self-Reported Workload	m > v		823	p = **.002***
Task Completion Time		m < v[4]	39	p **< .001***
Task Completion Time			898	p = .344
Heart Rate	m > v		1205	p = **.033***
Heart Rate	m > v		1326	p = .056[1]
SCL			743	p = .456
SCL	m > v		1340	p = **.005***
Mean Velocity			695	p = .106
Mean Velocity	m < v[5]		590	p = **.003***
SD Velocity			754	p =.237
SD Velocity			1099	p =.696
Acceleration Events (=>.1g)	m > v		654	p = **.004***
Acceleration Events (=>.1g)	m > v		559	p = **.019***
SD Steering Wheel Angle			1081	p = .223
SD Steering Wheel Angle	m > v		1370	p = **.028***
Minor SWR			1064	p = .274
Minor SWR			976	p = .671
Major SWR	m > v		1636	p **< .001***
Major SWR	m > v		1675	p **< .001***

TEORT = Total-Eyes-Off-Road-Time (NHTSA metric); GTD = Glance-to-Device; *statistically significant

[1]As detailed in the text, the radio hard tuning tasks differed somewhat between Study 1 and Study 2, with the former following the CAMP format and the latter following the current NHTSA guidelines.

[2]The distribution of the V statistic for the Wilcoxon test is such that smaller and larger values are associated with the two ends of the probability distribution (and hence statistical significance).

[3]The P values reported in this table may vary slightly from the *Statistical Comparison of Selected Measures* section in the individual detailed measure reporting sections since those tests are based on paired comparisons derived from analyses considering a larger set of tasks.

[4]It is open to question as to whether a longer task completion time should automatically be considered in the "negative" column for the voice interface.

[5]See discussion on a reduction in mean speed as a compensatory response to workload.

Table 6 summarizes the pattern of findings for non-glance based metrics comparing voice-command based engagement with the radio "hard" tuning task with use of the traditional visual-manual DVI. The pattern of findings across Study 1 and 2 are consistent in directionality. Self-reported workload in both studies was significantly lower for the voice-interface, as were the number of acceleration events (of 0.1g or greater), and the number of major steering wheel reversals. Variability (SD) of steering wheel angle did not show a significant difference in by interface type in Study 1, but was lower at a statistically significant level ($p = .028$) in Study 2. Mean velocity during the task trended lower during manual radio tuning in Study 1, but reached statistical significance in Study 2. Since slowing of vehicle speed is often interpreted as a compensatory behavior on the part of the driver to attempt to manage workload, a lower driving speed during manual radio tuning can be taken as evidence of greater demand on certain aspects of the driver's resources, particularly in light of the number of other measures such as self-reported workload ratings.

In terms of physiological indicators of arousal, heart rate was lower during engagement with the voice interface in Study 1 ($p = .033$) and trended in the same direction in Study 2 ($p = .056$). If the heart rate findings in Study 1 were used to generate a directional hypothesis for testing significance, it could be argued that it would be appropriate to apply a one-tailed test, in which case heart rate difference between the two interface conditions could be considered statistically significant. In terms of skin conductance level (SCL), while this index of arousal was not significantly different between the two conditions in Study 1, arousal did appear lower during the period the voice-interface was utilized than during manual tuning ($p = .005$).

As will be considered in more detail in the discussion, the NHTSA style radio tuning task used in Study 2 markedly reduced the task time differences between interfaces. The mean value for completing the "hard" radio tuning task in Study 1 was 48.1s using the voice interface vs. 24.9s using the manual ($p < .001$). In Study 2, where the radio was already active, the mean task completion times were not significantly different ($p = .344$), with the voice interface value of 27.4s actually being nominally shorter than the 28.4s using the manual interface.

Summary Findings Comparing Voice Nav. Entry and Manual Radio Tuning

A summary table (Table 38) is provided in the *Glance Analysis* section of this report that considers the visual demands associated with voice-based full destination address entry into the navigation system based on NHTSA guidelines. The tables below provide an alternate but conceptually related approach to looking at the demands and impact on the driver by comparing outcome measures for the navigation entry (Nav E) vs. the manual radio "hard" tuning (mRH) task. The "+ Voice Nav E" column indicates measures for which the voice-command method might reasonably be interpreted as being less demanding or having a smaller impact on that variable than is seen when engaging in the manual radio tuning reference task. Descriptive statistics are provided in the individual measure sections that follow.

Table 7: Glance Metrics Comparisons for Voice-Based Entry of a Full Destination Address into a Navigation System and Manual Radio Hard Tuning[1] (Wilcoxon tests[2]) across Studies 1 (Shaded) & 2 (Unshaded).

Measure	+ Voice Nav E	- Voice Nav E	V statistic[2]	p value
TEORT Mean Glance Duration	mRH > Nav E		87	p **< .001***
TEORT Mean Glance Duration	mRH > Nav E		159	p **< .001***
TEORT % Glances > 2s	mRH > Nav E		64	p **< .001***
TEORT % Glances > 2s	mRH > Nav E		60	p **< .001***
TEORT Total Glance Time		mRH < Nav E	1470	p **< .001***
TEORT Total Glance Time		mRH < Nav E	1859	p **< .001***
TEORT Number of Glances		mRH < Nav E	1430	p **< .001***
TEORT Number of Glances		mRH < Nav E	1993	p **< .001***
GTD Mean Glance Duration	mRH > Nav E		232	p **< .001***
GTD Mean Glance Duration	mRH > Nav E		310	p **< .001***
GTD % Glances > 2s	mRH > Nav E		76	p **< .001***
GTD % Glances > 2s	mRH > Nav E		79	p **< .001***
GTD Total Glance Time		mRH < Nav E	1165	p **< .001***
GTD Total Glance Time		mRH < Nav E	1440	p **< .001***
GTD Number of Glances		mRH < Nav E	1137	p **< .001***
GTD Number of Glances		mRH < Nav E	1565	p **< .001***

TEORT = Total-Eyes-Off-Road-Time (NHTSA metric); GTD = Glance-to-Device; *statistically significant

[1]As detailed in the text, the radio hard tuning tasks differed somewhat between Study 1 and Study 2, with the former following the CAMP format and the latter following the current NHTSA guidelines.

[2]The distribution of the V statistic for the Wilcoxon test is such that smaller and larger values are associated with the two ends of the probability distribution (and hence statistical significance).

It can be readily observed in Table 7 that the measures *mean glance duration* and the *percentage of glances longer than two seconds* are both lower during voice-based destination entry than during the manual radio tuning reference task. This pattern holds for both the EFOR and the GTD metrics and is observed in the data from both Study 1 and Study 2. It is also clear that *total glance time* and the *total number of glances* are consistently higher for the voice-based interface considered in this study relative to that observed during manual radio tuning.

Table 8: Other Comparisons for Voice-Based Entry of a Full Destination Address (Nav E) into a Navigation System and Manual Radio Hard (mRH) Tuning[1] (Wilcoxon tests[2]) across Studies 1 (Shaded) & 2 (Unshaded).

Measure	+ Voice Nav E	- Voice Nav E	V statistic[2]	p value[3]
Self-Reported Workload			**402**	**p = .275**
Self-Reported Workload			932	p = .173
Task Completion Time		mRH < Nav E	1830	p **< .001***
Task Completion Time		mRH < Nav E	2080	p **< .001***
Heart Rate	mRH > Nav E		506	p **= .003***
Heart Rate	mRH > Nav E		706	p = **.026***
SCL			679	p = .884
SCL			901	p **=** .752
Mean Velocity			1025	p = .420
Mean Velocity	mRH < Nav E[4]		1416	p = **.012***
SD Velocity		mRH < Nav E	1636	p **< .001***
SD Velocity		mRH < Nav E	1804	p **< .001***
Acceleration Events (=>.1g)			705	p = .589
Acceleration Events (=>.1g)			728	p = .904
SD Steering Wheel Angle		mRH < Nav E	1564	p **< .001***
SD Steering Wheel Angle		mRH < Nav E	1744	p **< .001***
Minor SWR			876	p = .777
Minor SWR			1195	p = .302
Major SWR	mRH > Nav E		113	p **< .001***
Major SWR	mRH > Nav E		511	p **< .001***

TEORT = Total-Eyes-Off-Road-Time (NHTSA metric); GTD = Glance-to-Device; *statistically significant

[1]As detailed in the text, the radio hard tuning tasks differed somewhat between Study 1 and Study 2, with the former following the CAMP format and the latter following the current NHTSA guidelines.

[2]The distribution of the V statistic for the Wilcoxon test is such that smaller and larger values are associated with the two ends of the probability distribution (and hence statistical significance).

[3]The P values reported in this table may vary slightly from the *Statistical Comparison of Selected Measures* section in the individual detailed measure reporting sections since those tests are based on paired comparisons derived from analyses considering a larger set of tasks.

[4]See discussion on a reduction in mean speed as a compensatory response to workload.

Self-reported workload was nominally higher for full destination address entry into the navigation system using the voice interface than for the “hard” manual radio tuning task in both studies. However, this difference was not statistically significant in either Study 1 or Study 2 ($p = .275$ and $p = .173$ respectively). Task completion time clearly differentiates the tasks ($p < .001$), with destination entry taking significantly longer at nearly two minutes (113s) to complete during Study 2 vs. 27.4s for manual radio tuning. Overall patterning will be considered further in the Discussion; however, it is clear that there are measures under which use of the voice-involved destination entry task appears to place no more demand on the driver than the manual radio tuning reference task, and others in which interpretation of the variable may be more open for review. (For example, to what extent are the standard deviation of velocity and steering wheel angle variables impacted by the longer duration of the destination entry task?)

Summary Findings Considering Age

Consistent with Study 1, age impacted self-reported workload, task completion time, mean velocity, number of acceleration events, TEORT, and measures of the percentage of glances greater than 2 seconds. Speciffically, participants in the two older age groups showed higher mean self-reported workload levels than participants in the two younger age groups (see Figure 15, p. 56); interestingly, there was a trend in some tasks in which the 25-39 year old group showed the lowest self-reported workload levels (i.e. nominally lower than that reported by 20-24 year olds). Task completion time generally increased with age, with the 55+ age group showing the longest mean completion times across tasks. As has been noted in numerous studies as an age relationship, the oldest participant group (55+) drove slower than the younger participants.

Table 9: Statistical Significance for Main Effects of Age on Each of the Primary Outcome Measures

Measure	F statistic	p value
Self-Reported Workload	3.34	p = **.026***
Task Completion Time	7.73	p < **.001***
Heart Rate	0.20	p = .894
SCL	0.41	p = .744
Mean Velocity	3.79	p = **.015***
SD Velocity	0.21	p = .891
Acceleration Events (=>.1g)	3.23	p = **.029***
SD Steering Wheel Angle	0.27	p = .844
Minor SWR	2.25	p = .092
Major SWR	0.20	p = .895
TEORT Mean Glance Duration	2.37	p = .081
TEORT % Glances > 2s	2.67	p = .056
TEORT Total Glance Time	2.90	p = **.043***
TEORT Number of Glances	2.12	p = .108
GTD Mean Glance Duration	2.07	p = .114
GTD % Glances > 2s	2.94	p = **.041***
GTD Total Glance Time	0.94	p = .427
GTD Number of Glances	0.50	p = .683

TEORT = Total-Eyes-Off-Road-Time (NHTSA metric); GTD = Glance-to-Device; *statistically significant

While older participants drove slower, younger participants showed a greater number of acceleration events in the 1g or greater range (see Figure 55, p. 94). This effect was driven in part by an age by gender interaction in which younger males as a group generally showed higher frequency counts; for some tasks this was most evident in the 25-39 year old male samples.

Age effects for glance metrics were statistically more robust in Study 1, most likely due to the fact that Study 1 considered relatively young (20-29 year olds) against relatively older participants (60-69 years), whereas Study 2 considers an age distribution across four continuous age groups (20-24, 25-39, 40-54, and 55+). Nonetheless, the patterning is consistent in TEORT glance metrics; mean glance duration, the percentage of long duration glances, and total glance time all trend toward being longer as age increases. This effect is statistically significant for total glance time off road using the TEORT measure. Mean glance duration and the percentage of long duration glances both show a significant age by gender interaction, with the two older male groups driving the longer glance values.

Summary Findings Considering Gender

In Study 1, when considering the GTD metric, males showed a statistically higher percentage of long duration glances (> 2 seconds) and a trend toward longer mean single glance durations. This pattern was much more robust in the current study as can be observed in Table 10 below. Both the GTD and TEORT metrics showed statistically significant main effects for mean single glance duration and percentage of long duration (> 2 seconds) glances. In both instances, males showed higher values. As detailed in the glance analysis section (see Figure 80p. 123 and Figure 87 p. 130), this effect is driven by males in the two older age categories (40+) in an age by gender interaction (older males diverging from older females and showing longer mean single glance durations and a higher percentage of long duration glances).

Table 10: Statistical Significance for Main Effects of Gender on Each of the Primary Outcome Measures

Measure	F statistic	p value
Self-Reported Workload	0.44	p = .512
Task Completion Time	0.54	p = .464
Heart Rate (% change)	0.42	p = .518
SCL (% change)	1.56	p = .218
Mean Velocity	1.06	p = .307
SD Velocity	1.48	p = .228
Acceleration Events (=>.1g)	1.67	p = .201
SD Steering Wheel Angle	3.36	p = .072
Minor SWR	0.64	p = .427
Major SWR	0	p = .956
TEORT Mean Glance Duration	8.65	p = **.005***
TEORT % Glances > 2s	20.03	p < **.001***
TEORT Total Glance Time	0.58	p = .449
TEORT Number of Glances	0.22	p = .645
GTD Mean Glance Duration	7.05	p = **.010***
GTD % Glances > 2s	20.38	p < **.001***
GTD Total Glance Time	2.15	p = .148
GTD Number of Glances	0.49	p = .489

TEORT = Total-Eyes-Off-Road-Time (NHTSA metric); GTD = Glance-to-Device; *statistically significant

It may be worth noting again here that the video quality for manual glance analysis was higher in this second study since it was designed from the initiation of the study to be used for glance analysis in place of automated eye tracking. Good quality video for manual glance coding was a requirement for cases to be included in Study 2. Consequently, the glance analysis data presented here is based on 64 cases equally balanced by age group and gender while glance analysis data presented for Study 1 was based on a somewhat smaller sample of 53 usable cases (for most glance measures) out of 60 participants. This further increases the confidence in the gender patterning reported here.

The only other gender effect appearing in the data from Study 1 was the finding that the female sample showed a somewhat higher percentage change in SCL during secondary tasks than males. As we noted in the technical report for Study 1, this was not a pattern we had observed in other datasets and indicated that it should be interpreted cautiously. As can be observed in Table 10, no significant effect of gender on percentage change in SCL during secondary tasks appeared in Study 2.

Self-Reported Workload

At the completion of the drive, participants rated how much workload they felt was involved in trying to do each task. A 0 (low) to 10 (high) scale that allowed for half point resolution was employed. (See Appendix A for additional detail.)

Table 11: Mean (and standard deviation) of self-reported workload ratings by training type.

	Structured Training	Self-Guided Training	(combined)
Nav Entry	3.97 (2.4)	3.05 (2.5)	3.51 (2.5)
Nav Cancel	4.37 (2.5)	3.05 (2.7)	3.70 (2.7)
Radio Manual Easy	2.23 (2.5)	1.59 (1.5)	1.91 (2.1)
Radio Manual Hard	3.25 (3.0)	2.73 (1.7)	2.99 (2.4)
Radio Voice Easy	2.45 (2.2)	1.62 (1.6)	2.04 (2.0)
Radio Voice Hard	2.14 (2.1)	1.81 (1.8)	1.98 (1.9)

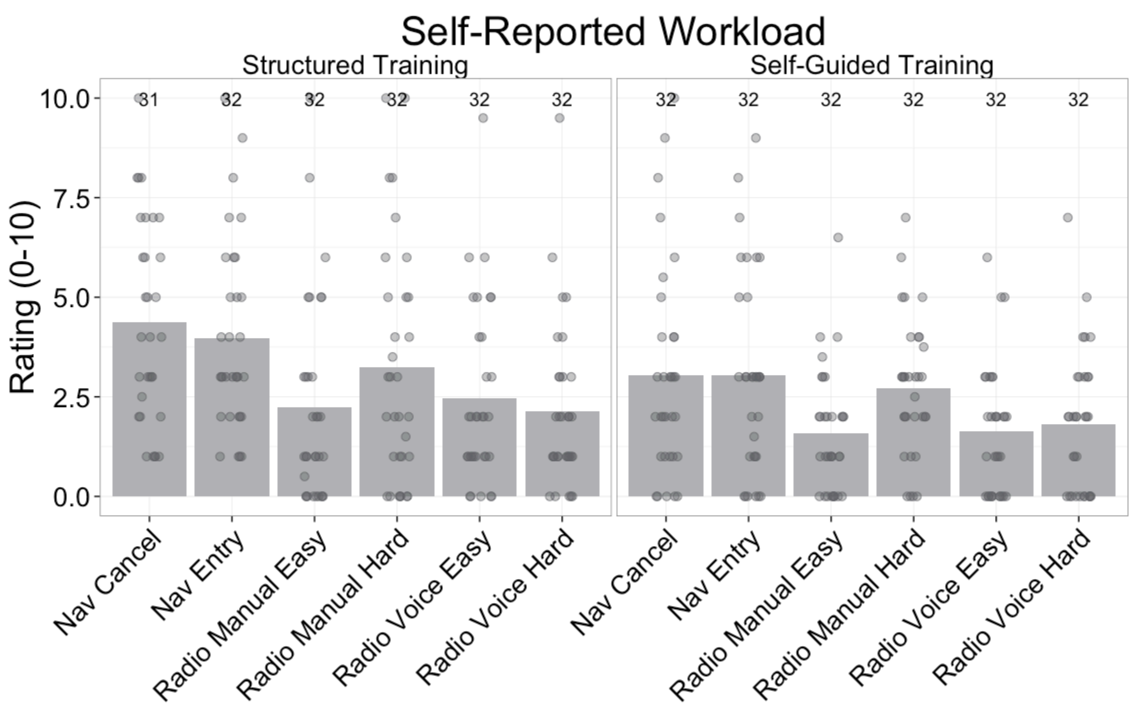


Figure 10: Self-reported workload ratings by task and training condition with solid circles the distribution of ratings by individual participants.

In contrast with pre-study expectations, workload ratings were nominally lower in the self-guided training condition, though given the high between-subject variability, this difference is not statistically significant (p = .109). Thus, the highly structured guidance provided in the directed training condition did not provide any clear advantage across the sample as a whole in terms of self-reported workload ratings after engaging in the tasks under actual driving conditions. Since the effect of training type was not significant, the figures below combine ratings across the training groups.

As can be seen in Figure 10, in which the two training conditions are presented separately, the Navigation Entry and Cancel tasks were consistently rated as the most difficult, the Radio Manual (Hard) task next, and the remaining tasks following with nominal differences between them. Thus, in addition to there being no statistically significant difference between the conditions on the basis of a repeated measures analysis of variance, the relative consistency of patterning argues for the appropriateness of collapsing across the training conditions and combining them into a larger overall sample of 64 participants for assessing the individual tasks.

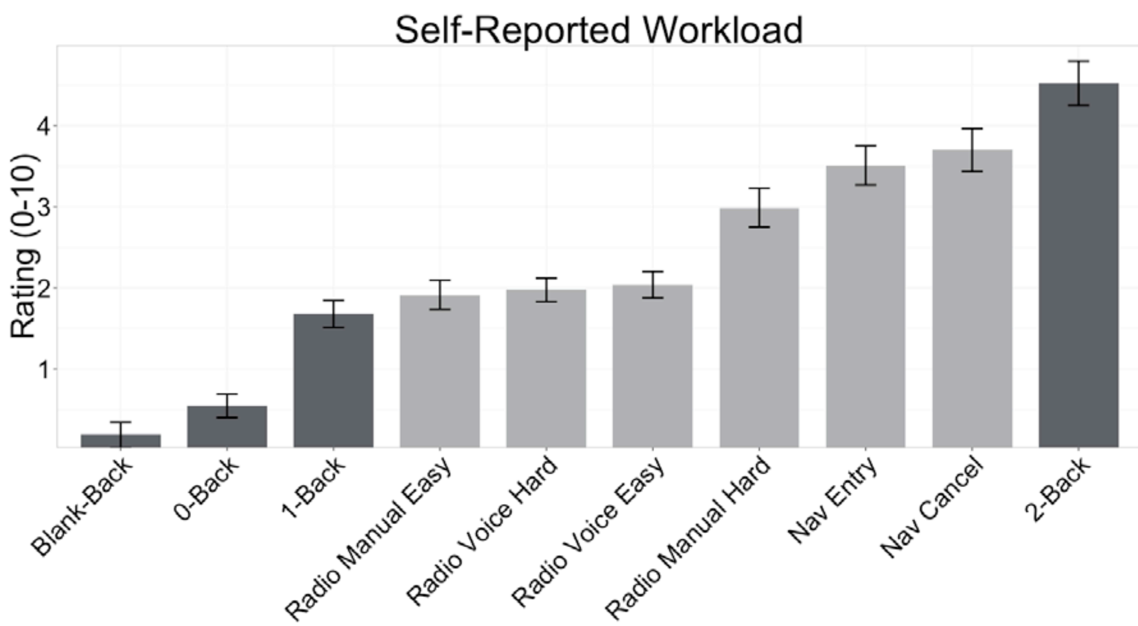


Figure 11: Mean workload ratings collapsed across training type and displaying adjusted standard errors. Ratings for the four levels of the n-back task are presented as dark gray bars.

When tasks are ordered by self-reported workload, ratings for the DVI tasks all fall between the ratings for the 1-back and 2-back. This is largely consistent with findings for Study 1 with the notable exception of the navigation cancelation task (Nav Cancel) which was nominally rated as involving the most workload of any DVI task in Study 2.

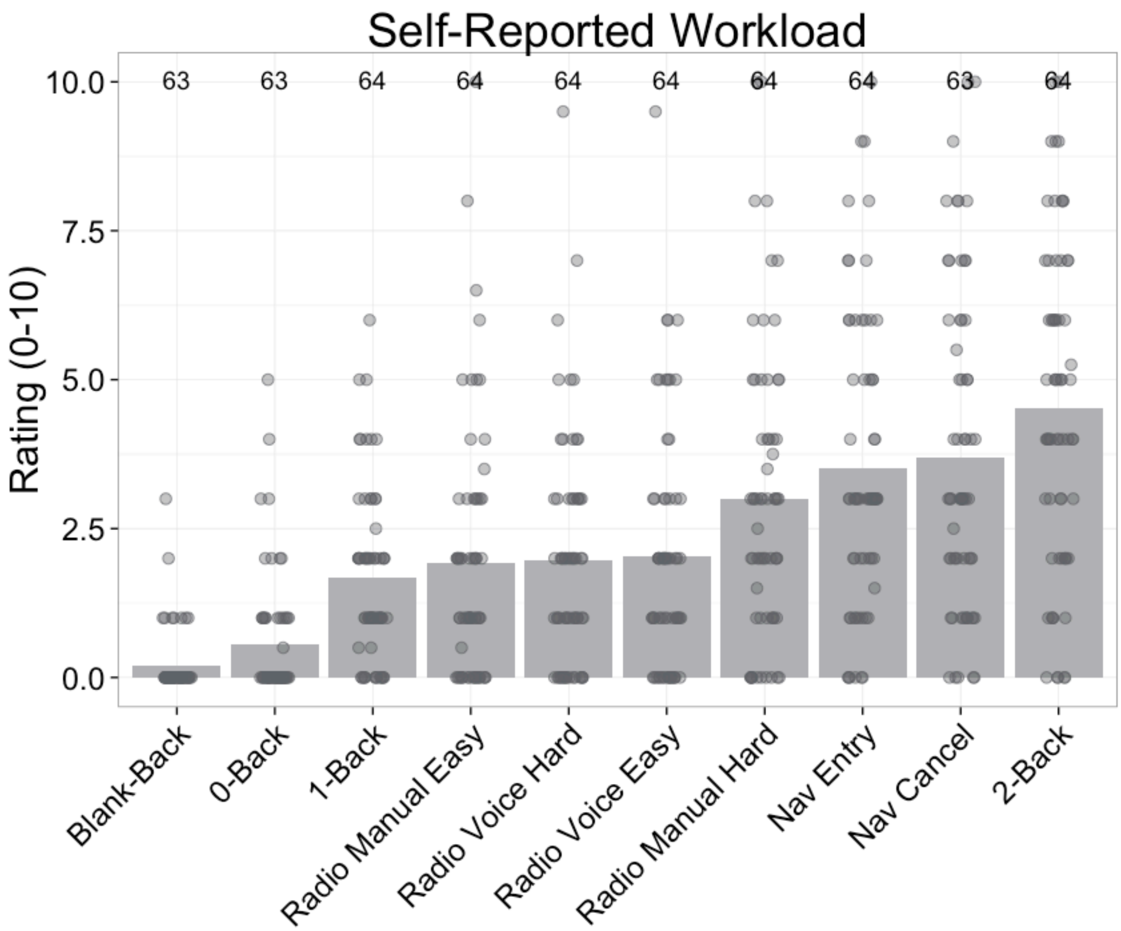


Figure 12: Mean values with solid circles showing the distribution of ratings by individual participants.

It is apparent in the plot of individual workload ratings that participants vary widely in the absolute values that they assign to tasks (Figure 12). However, the relative values assigned across tasks appear to have a great deal of face validity, with the n-back tasks lining-up in order from blank-back through the 0-, 1- and 2-back levels.

Comparison to Study 1

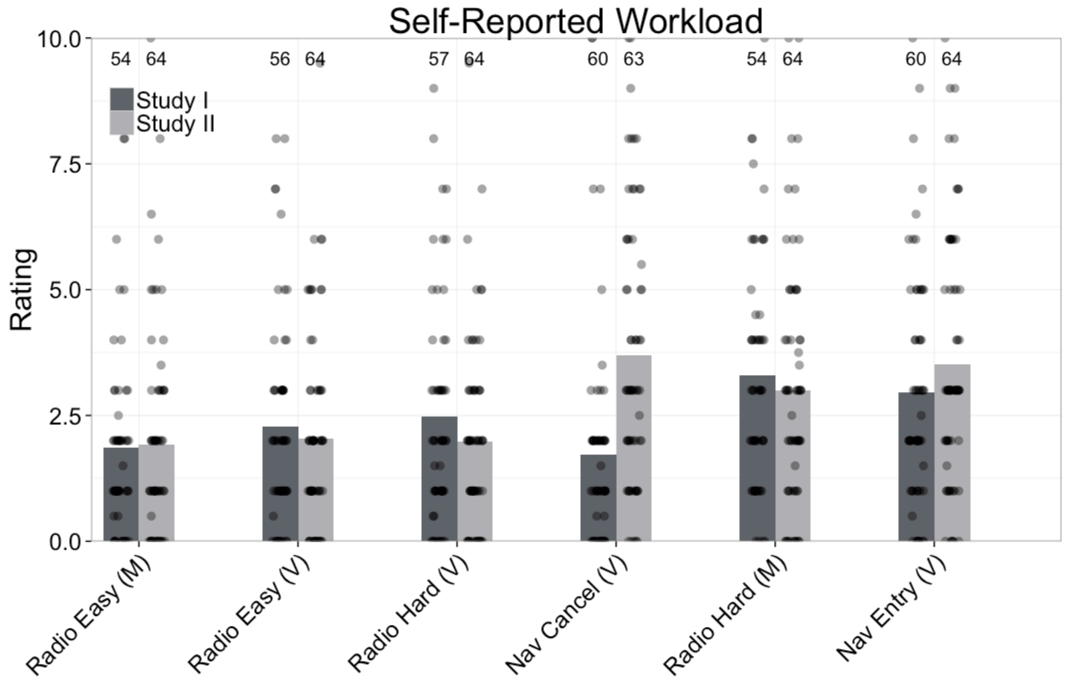


Figure 13: Comparison of self-reported workload in Study 1 and Study 2.

Participants in the present study gave the Navigation Cancel task an average workload rating of 3.70, whereas in Study 1, this task was rated much lower, with a mean rating of 1.69. This is likely due to fact that the task prompt in Study 1 explicitly provided participants with the command required to cancel the task (i.e. "*Your task is to cancel the route using the command '**Navigation Cancel Route**'.*"). The task prompt was changed in this study to "*Your task is to cancel the route you entered.*", which meant that participants who went through the structured training had to recall the proper form of the command or rely solely on the cues provided by the system like the self-guided training group. This change was made in the current study so that the self-trained group was not provided with explicit instructions in how to complete the task. It appears from self-report that, as a group, the structured training participants had at least as much difficulty recalling the exact structure of this command as the self-guided participants (Figure 10).

Manual tuning of the radio (Radio Manual Hard) was again rated around 3 (Study 1: M 3.31; SD 2.3 - Study 2: M 2.99; SD 2.4), which is significantly higher than the rating for voice-based radio tuning (Radio Voice Hard) (Study 1: M 2.48; SD 2.2 - Study 2: M 1.98; SD1.9). As detailed in the next section, this difference between manual and voice-based radio tuning is statistically significant. Navigation system address entry (Nav Entry) was rated nominally higher on workload than manual radio tuning (Radio Manual Hard), which is consistent with Study 1; however, as detailed in the next section, this difference was not statistically significant.

Statistical Comparison of Selected Tasks

Throughout the results presentation in this report, statistical assessments of three DVI tasks are considered: voice-command based address entry into the navigation system (Nav Entry), multi-step manual tuning of the radio (Radio Manual Hard), and voice-command based engagement with the same task (Radio Voice Hard). Where appropriate, data on the 1-back task is also presented.

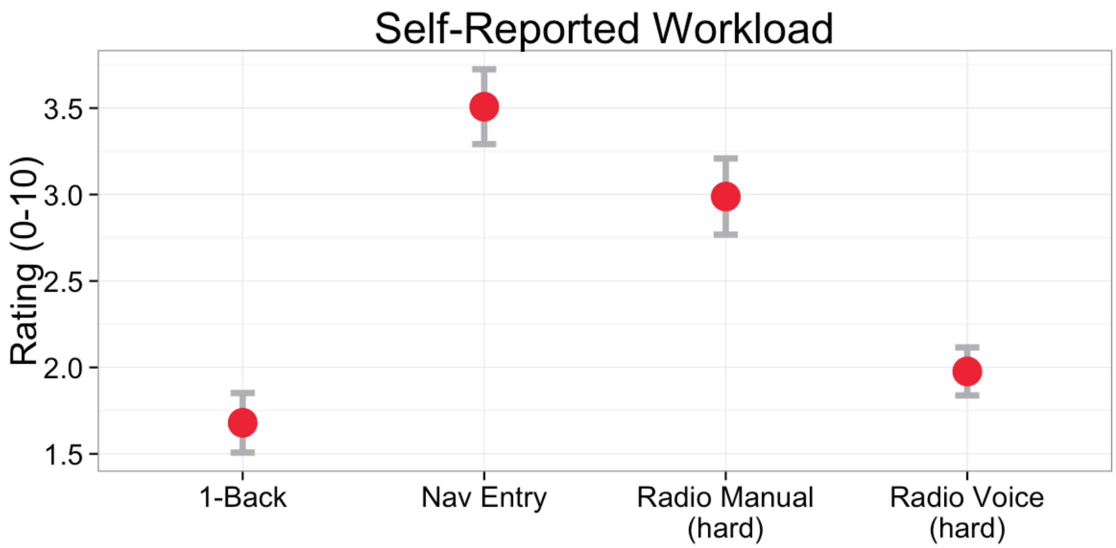


Figure 14: Workload measurements across selected reference tasks.

Workload differed significantly across the four reference tasks ($X^2 = 32.1$, $p < .001$, Friedman test). The Navigation Entry task was rated as having the highest level of demand, statistically similar to the Radio Manual Hard task ($p = .17$, NS). The 1-back and Radio Voice tasks were rated as having lower workload. The Radio Voice task was rated as having a lower workload than the equivalent Radio Manual task ($p = .002$).

This overall pattern of ratings is similar to what was found in Study 1, with workload ratings for the DVI reference tasks falling in-between the 1-back and 2-back levels, and voice tuning of a specified radio station being rated as involving less workload than manual radio tuning at a statistically significant level. Workload ratings of navigation entry and manual radio tuning did not show a statistically significant difference from each other in either study.

Descriptive Statistics by Task with Age & Gender Breakdowns

Descriptive statistics (mean and standard deviation) for workload ratings split by age and gender are provided in Table 5 below and presented graphically in Figure 15 and Figure 16 on the following page. This same presentation format is continued throughout this report.

Table 12: Mean and (SD) of workload ratings by task, broken down by age group and gender.

Combined	20-24	25-39	40-54	55+	(all)
Nav Entry	3.56 (2.8)	3.00 (2.0)	3.44 (2.1)	4.03 (2.9)	3.51 (2.5)
Nav Cancel	3.38 (3.1)	2.43 (2.1)	4.22 (2.4)	4.69 (2.6)	3.70 (2.7)
Radio Manual Easy	1.38 (2.1)	1.31 (1.4)	2.56 (1.8)	2.41 (2.8)	1.91 (2.1)
Radio Manual Hard	2.88 (2.6)	1.78 (1.6)	3.28 (1.9)	4.02 (3.0)	2.99 (2.4)
Radio Voice Easy	1.12 (1.0)	1.75 (1.8)	2.44 (2.0)	2.84 (2.6)	2.04 (2.0)
Radio Voice Hard	1.19 (1.2)	1.25 (1.2)	2.38 (2.1)	3.09 (2.3)	1.98 (1.9)
Blank-Back	0.06 (0.2)	0.31 (0.6)	0.20 (0.4)	0.19 (0.8)	0.19 (0.5)
0-Back	0.69 (1.4)	0.31 (0.6)	0.70 (0.9)	0.50 (1.2)	0.55 (1.0)
1-Back	1.50 (1.5)	1.91 (1.6)	1.91 (1.3)	1.41 (1.3)	1.68 (1.4)
2-Back	3.94 (2.4)	4.08 (2.7)	4.81 (2.1)	5.25 (3.5)	4.52 (2.7)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	3.50 (3.4)	2.75 (2.4)	3.75 (2.4)	4.25 (3.1)	3.56 (2.8)
Nav Cancel	4.25 (3.7)	1.93 (1.2)	5.31 (1.8)	5.12 (3.1)	4.23 (2.9)
Radio Manual Easy	1.12 (1.4)	1.62 (1.8)	2.62 (1.9)	2.88 (3.1)	2.06 (2.2)
Radio Manual Hard	2.88 (2.7)	1.69 (1.8)	3.56 (2.0)	5.53 (2.9)	3.41 (2.7)
Radio Voice Easy	0.88 (1.0)	2.00 (2.0)	2.38 (2.0)	2.69 (3.2)	1.98 (2.2)
Radio Voice Hard	1.00 (1.2)	1.50 (1.4)	2.00 (2.2)	3.44 (3.0)	1.98 (2.2)
Blank-Back	0.00 (0.0)	0.50 (0.8)	0.29 (0.5)	0.00 (0.0)	0.19 (0.5)
0-Back	0.62 (0.9)	0.38 (0.7)	0.71 (0.8)	0.12 (0.4)	0.45 (0.7)
1-Back	1.38 (1.5)	2.12 (2.2)	1.56 (1.5)	1.06 (1.0)	1.53 (1.6)
2-Back	4.12 (2.5)	3.62 (2.4)	3.62 (2.2)	5.50 (4.4)	4.22 (3.0)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	3.62 (2.3)	3.25 (1.8)	3.12 (1.9)	3.81 (3.0)	3.45 (2.2)
Nav Cancel	2.50 (2.3)	2.88 (2.7)	3.12 (2.5)	4.25 (2.0)	3.19 (2.4)
Radio Manual Easy	1.62 (2.7)	1.00 (0.9)	2.50 (1.8)	1.94 (2.5)	1.77 (2.1)
Radio Manual Hard	2.88 (2.6)	1.88 (1.6)	3.00 (1.9)	2.50 (2.4)	2.56 (2.1)
Radio Voice Easy	1.38 (1.1)	1.50 (1.7)	2.50 (2.1)	3.00 (2.0)	2.09 (1.8)
Radio Voice Hard	1.38 (1.3)	1.00 (0.9)	2.75 (2.1)	2.75 (1.5)	1.97 (1.6)
Blank-Back	0.12 (0.4)	0.12 (0.4)	0.12 (0.4)	0.38 (1.1)	0.19 (0.6)
0-Back	0.75 (1.8)	0.25 (0.5)	0.69 (1.0)	0.88 (1.6)	0.64 (1.3)
1-Back	1.62 (1.6)	1.69 (1.0)	2.25 (1.0)	1.75 (1.6)	1.83 (1.3)
2-Back	3.75 (2.4)	4.53 (3.0)	6.00 (1.1)	5.00 (2.6)	4.82 (2.4)

Visualization of the key DVI tasks by age and gender demographics is provided in plots such as the figures below throughout this report.

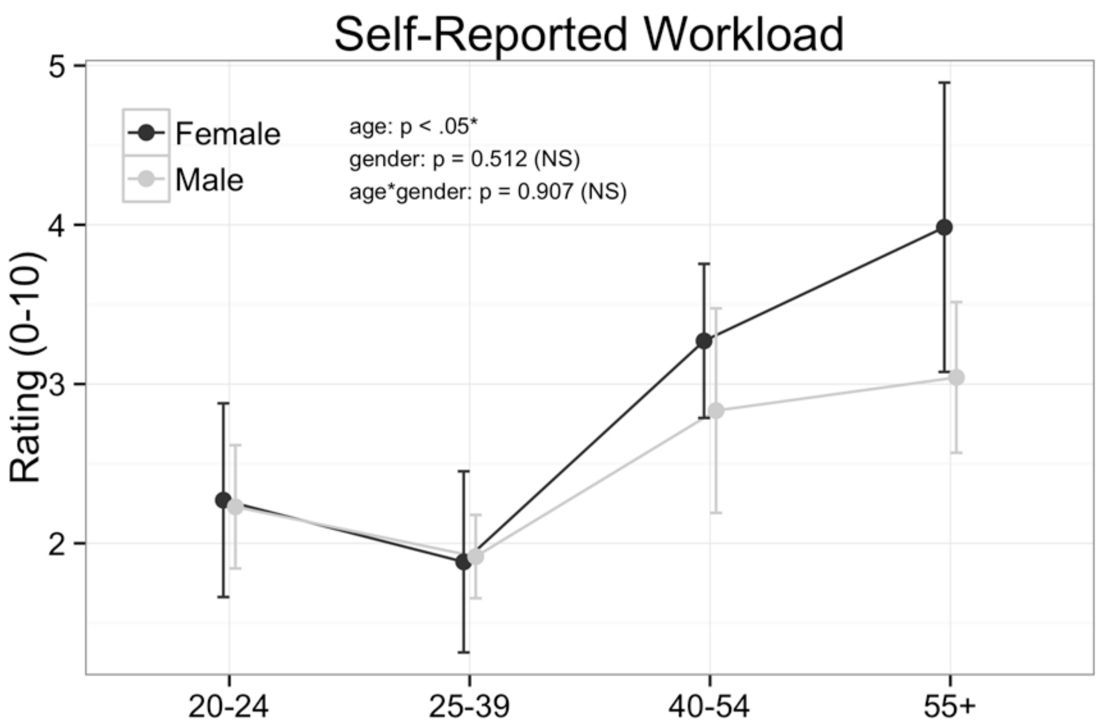


Figure 15: Workload ratings across all vehicle interface tasks (Nav Entry, Nav Cancel, and all Radio tasks) by age group and gender.

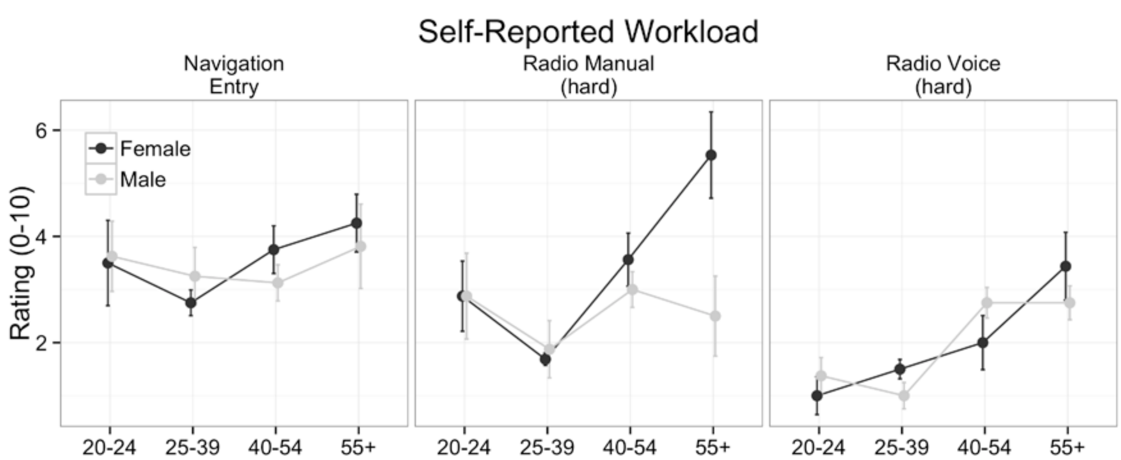


Figure 16: Workload ratings by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Mean self-reported workload ratings collapsed across all DVI tasks (i.e. n-back ratings are excluded) by age and gender are presented in Figure 15. Overall, there is a main effect of age ($p < .05$) with ratings tending to increase with age, although it can be seen that overall mean workload does appear nominally the lowest in the 25-39 age group. While there was no overall main effect of gender across tasks ($p = 0.512$) and no by age by gender interaction ($p = 0.907$), potentially interesting sub-patterns are suggested in the selected age by gender plots for the three selected tasks that are broken out for consideration throughout this report (Navigation Entry, Manual (Hard) Tuning of a Radio Station, and Voice (Hard) Tuning of a Radio Station) as seen in Figure 16.

Task Completion Time

Task completion time represents another way of evaluating the demand and potential distraction associated with a task. Figure 17 shows a plot of the total time in seconds from the start to the completion of each of the DVI tasks.

Table 13: Mean (and standard deviation) of task completion times by training type.

	Structured Training	Self-Guided Training	(combined)
Nav Entry	112.13 (37.5)	113.73 (24.3)	112.93 (31.4)
Nav Cancel	47.17 (23.7)	47.81 (34.5)	47.49 (29.4)
Radio Manual Easy	6.36 (3.3)	7.19 (5.4)	6.77 (4.5)
Radio Manual Hard	26.00 (16.6)	30.85 (18.5)	28.43 (17.6)
Radio Voice Easy	22.76 (8.3)	21.63 (6.1)	22.20 (7.3)
Radio Voice Hard	26.62 (10.5)	28.19 (7.6)	27.40 (9.1)

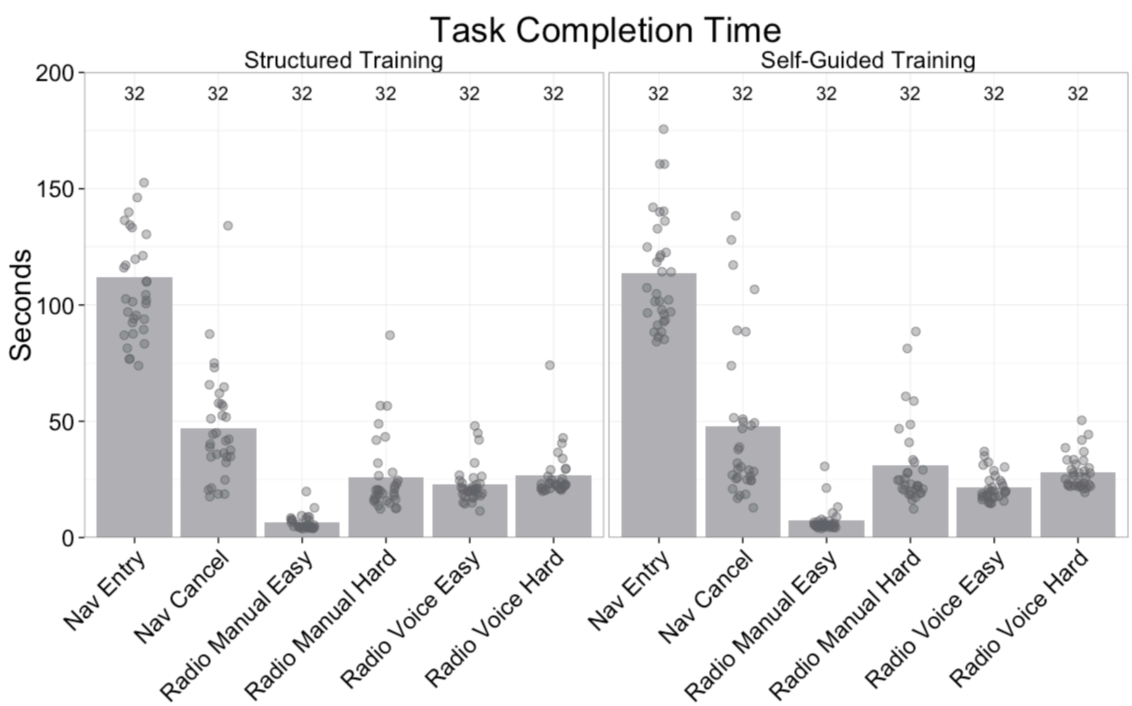


Figure 17: Task completion time by task and training condition.

Task completion times between the two training conditions were not significantly different ($p = .257$). It is evident that training type did not markedly impact efficiency in completing the tasks for the group as a whole. The degree of similarity between the two training conditions is remarkable; even relatively small fluctuations in completion time across task (as in the Radio Manual Hard, Radio Voice Easy, and Radio Voice Hard bars) are reflected in the patterning of the plots of mean values for both training conditions. The figures below present completion times by task collapsed across the entire sample of 64 participants (the n-back tasks are excluded due to the fixed trial length employed in those tasks).

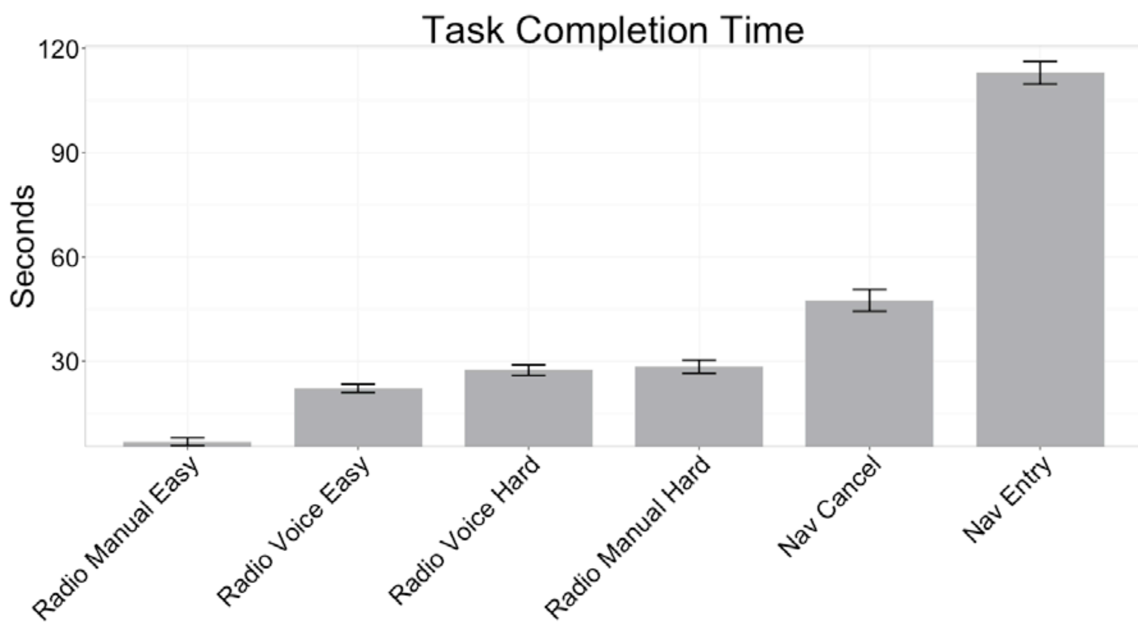


Figure 18: Means and adjusted standard errors collapsing data across both training groups.

Consistent with Study 1 (Figure 20), the Navigation Entry task required more time to complete than any other task, in all likelihood due to the large number of steps this task requires (Study 1: 110.67 seconds; SD 33.5 – Study 2: 112.93 seconds; SD 31.4). The consistency of results for total task time across the two studies is quite notable.

Consistent with the self-reported workload data reported previously, participants in the present study took longer to complete the Navigation Cancel task as compared to the previous study (47.5s vs. 26.2s, respectively), a statistically significant difference ($t_{(108.7)} = 4.8$, $p < .001$, t-test). As already discussed, this does not appear related to the training method (structured training vs. self-guided) as task completion time was practically identical in both conditions. Instead, this most likely reflects an increase in difficulty brought on by the lack of an explicit task prompt to participants reminding them of the specific command to use to cancel the route request in this study.

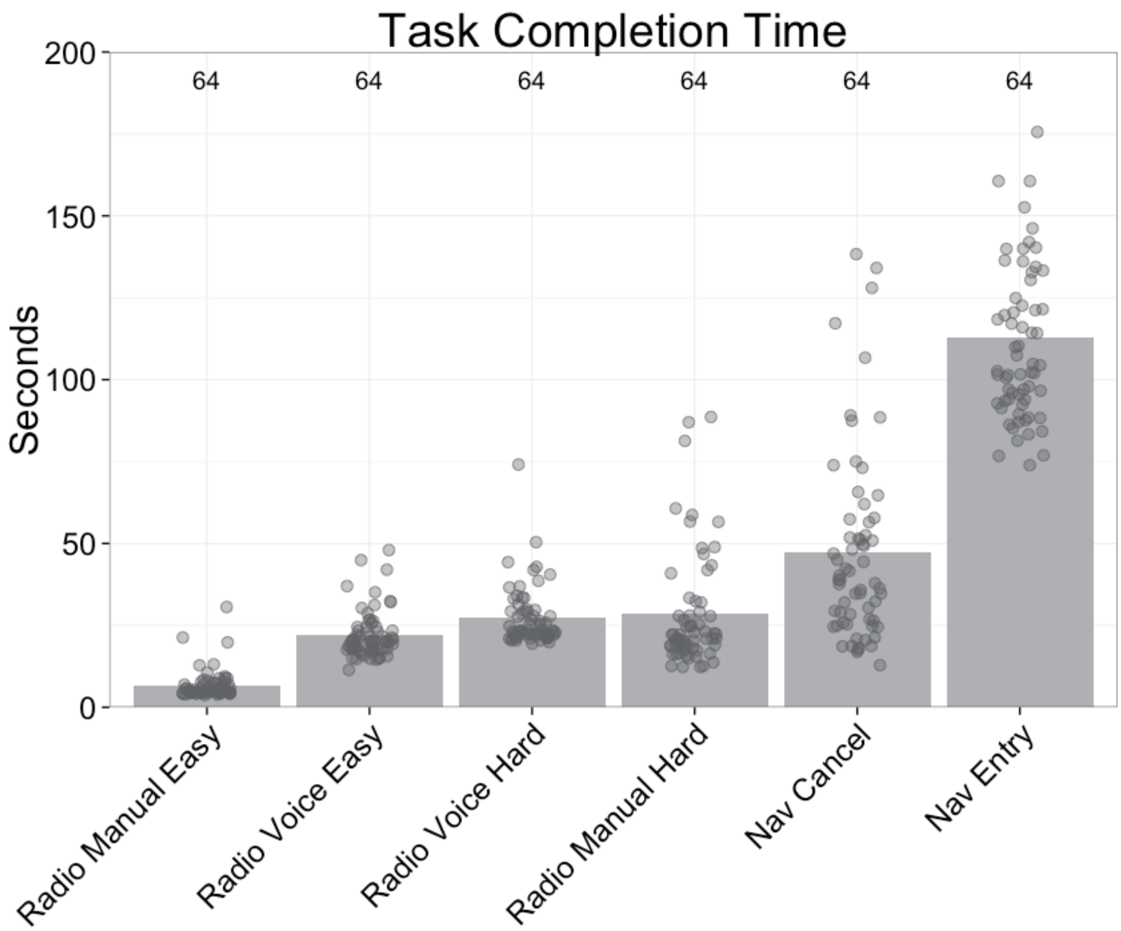


Figure 19: Task completion time collapsed across training conditions showing the distribution of individual participants.

It can be noted in the plot of individual task completion times for each task (Figure 19), that the Nav Cancel task shows a particularly range of values. This is most likely related to some individuals having clear recall of the specific command to use to cancel route navigation and other individuals having difficulty recalling the specific command sequence and either hesitating for an extended period of time or needing to try one or more attempts to get the correct phrasing within the allotted voice-response window.

Comparison to Study 1

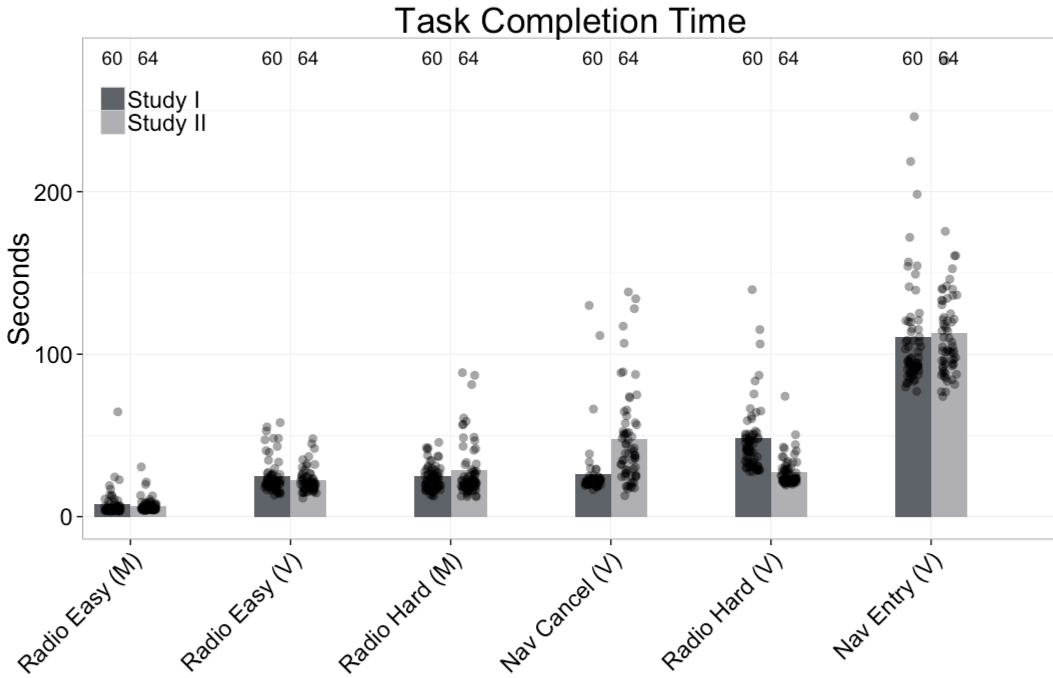


Figure 20: Comparison of task completion time in Study 1 and Study 2.

As can be observed in Figure 20, task completion times were quite consistent across studies except for the Nav Cancel and the voice version of the Radio Hard task. The Nav Cancel difference has already been discussed. The voice-command version of the radio "hard" task was made appreciably easier in Study 2 in that participants did not have to issue a command to turn the radio on and then issue a command to request a specific station. In Study 2, the radio was already on, thus reducing the task by one step. (See Appendix B for a detailed consideration of the reason for and the effective difference in the radio "hard" task in this study.)

Statistical Comparison of Selected Tasks

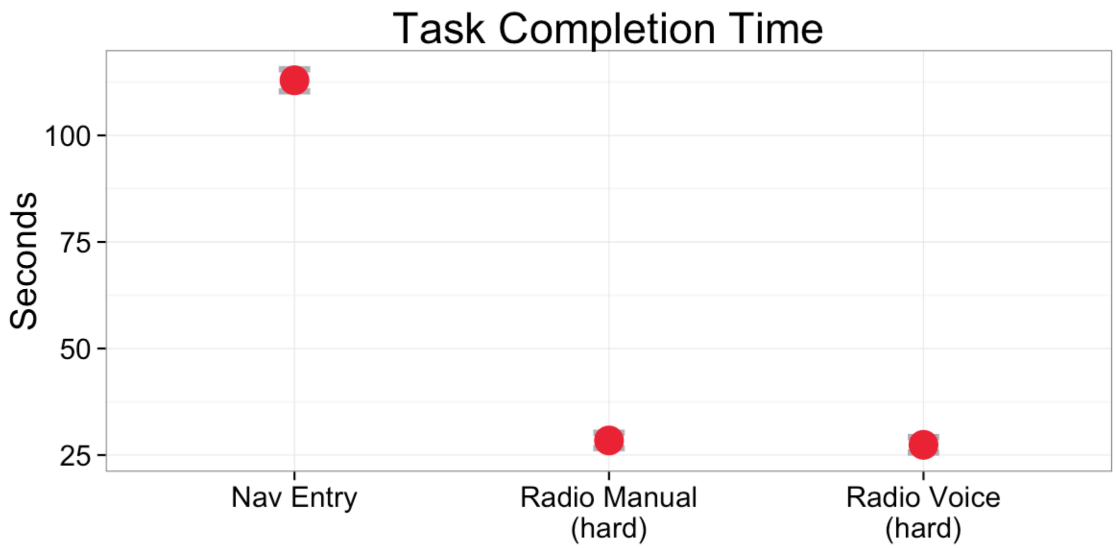


Figure 21: Task completion time across selected reference tasks.

Task completion time differed significantly across the three reference tasks ($X^2 = 97.5$, $p < .001$, Friedman test). The Navigation Entry task took nearly two minutes (113s) to complete, standing in stark contrast to either version of the Radio Hard tuning task, both of which were completed in under 30s and did not differ from each other ($p = .344$). These results are quite similar to those of Study 1, though in that case, the Radio Voice Hard task took somewhat longer to complete (48.1s), albeit still substantially faster than Navigation Entry.

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 14: Mean and (SD) of task completion times by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	103.12 (18.6)	103.34 (17.5)	110.91 (25.2)	134.35 (46.4)	112.93 (31.4)
Nav Cancel	50.78 (35.1)	39.95 (28.6)	43.68 (22.7)	55.56 (30.0)	47.49 (29.4)
Radio Manual Easy	5.96 (3.8)	5.61 (2.4)	7.01 (6.5)	8.51 (4.0)	6.77 (4.5)
Radio Manual Hard	23.13 (17.4)	26.41 (18.2)	25.37 (12.1)	38.80 (18.9)	28.43 (17.6)
Radio Voice Easy	18.75 (5.3)	20.64 (5.4)	22.41 (8.0)	26.99 (7.9)	22.20 (7.3)
Radio Voice Hard	26.02 (6.7)	28.61 (13.5)	27.16 (7.0)	27.82 (8.2)	27.40 (9.1)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	106.46 (21.9)	100.61 (17.9)	108.23 (31.0)	122.11 (21.7)	109.35 (23.8)
Nav Cancel	64.06 (42.2)	40.09 (39.2)	60.04 (18.6)	57.38 (26.7)	55.39 (32.8)
Radio Manual Easy	7.12 (5.3)	6.33 (3.1)	5.26 (1.6)	9.98 (5.2)	7.17 (4.3)
Radio Manual Hard	26.97 (24.4)	24.51 (9.4)	30.09 (15.3)	40.72 (19.1)	30.57 (18.1)
Radio Voice Easy	18.25 (2.4)	22.16 (6.7)	19.30 (2.6)	24.10 (4.4)	20.95 (4.8)
Radio Voice Hard	23.80 (4.4)	23.46 (2.4)	24.85 (3.9)	29.20 (10.2)	25.33 (6.2)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	99.79 (15.3)	106.07 (18.0)	113.60 (19.6)	146.59 (61.6)	116.51 (37.5)
Nav Cancel	37.49 (21.6)	39.81 (14.7)	27.32 (12.1)	53.74 (34.7)	39.59 (23.5)
Radio Manual Easy	4.80 (0.6)	4.89 (1.0)	8.76 (8.9)	7.05 (1.5)	6.37 (4.7)
Radio Manual Hard	19.29 (3.8)	28.30 (24.7)	20.65 (5.7)	36.88 (19.9)	26.28 (17.0)
Radio Voice Easy	19.25 (7.4)	19.12 (3.5)	25.52 (10.4)	29.88 (9.7)	23.44 (9.0)
Radio Voice Hard	28.24 (8.1)	33.76 (18.0)	29.47 (8.8)	26.44 (5.8)	29.48 (11.0)

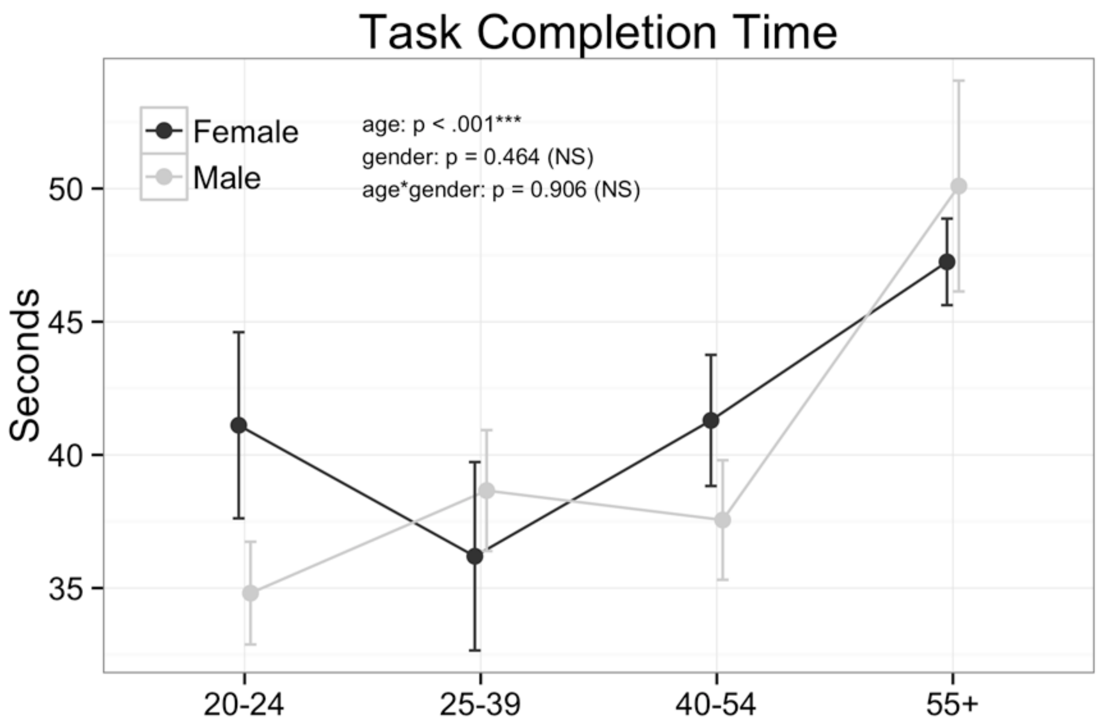


Figure 22: Task completion times across age groups and genders.

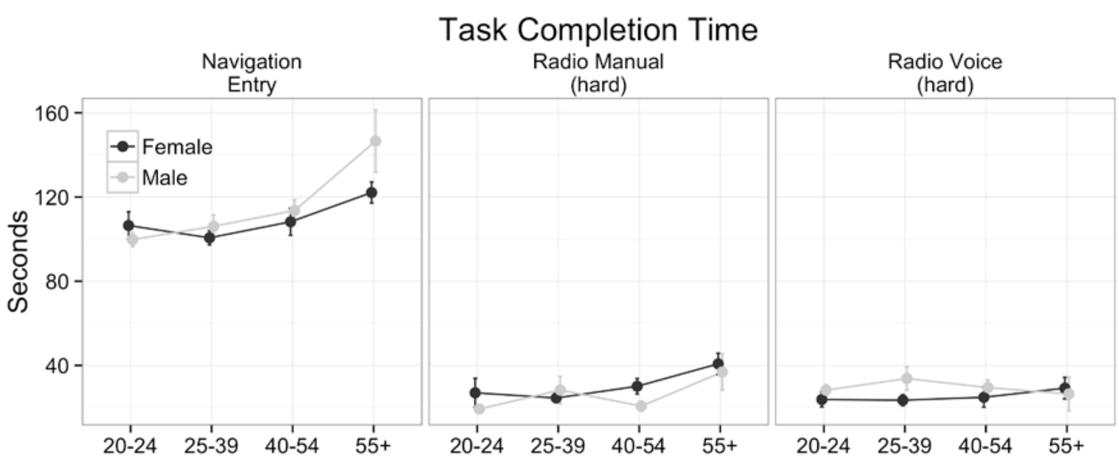


Figure 23: Task completion times by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Task completion time by age remains fairly stable until participants reach the oldest of the age groups, at which point they increase significantly ($p < .001$). This pattern is most apparent in the navigation entry and manual radio tuning task; it does not appear in the voice radio tuning task. There does not appear to be a substantial overall effect of gender on task completion time.

Physiological Measures

Heart Rate

Percent change in heart rate relative to the mean of each subject's two-minute baseline periods is shown in the figures and tables below. Changes in heart rate have been shown to be sensitive to changes in cognitive demand (Mehler et al., 2012).

Table 15: Mean (and standard deviation) of percent change in heart rate by training type.

	Structured Training	Self-Guided Training	(combined)
Nav Entry	1.55 (5.2)	2.22 (4.8)	1.89 (5.0)
Nav Cancel	1.16 (4.6)	2.33 (4.7)	1.74 (4.6)
Radio Manual Easy	2.05 (4.7)	1.09 (3.9)	1.57 (4.3)
Radio Manual Hard	3.47 (5.4)	2.78 (4.4)	3.13 (4.9)
Radio Voice Easy	1.59 (4.1)	1.18 (4.0)	1.38 (4.0)
Radio Voice Hard	1.97 (4.0)	2.17 (4.6)	2.07 (4.3)

Figure 24: Change in heart rate by task and training condition.

Heart rate change during tasks was not affected by training condition (p = .976). As with other measures, a similar gross patterning in changes in heart rate are evident in both training conditions, with the Radio Manual Hard task showing the greatest elevation in heart rate.

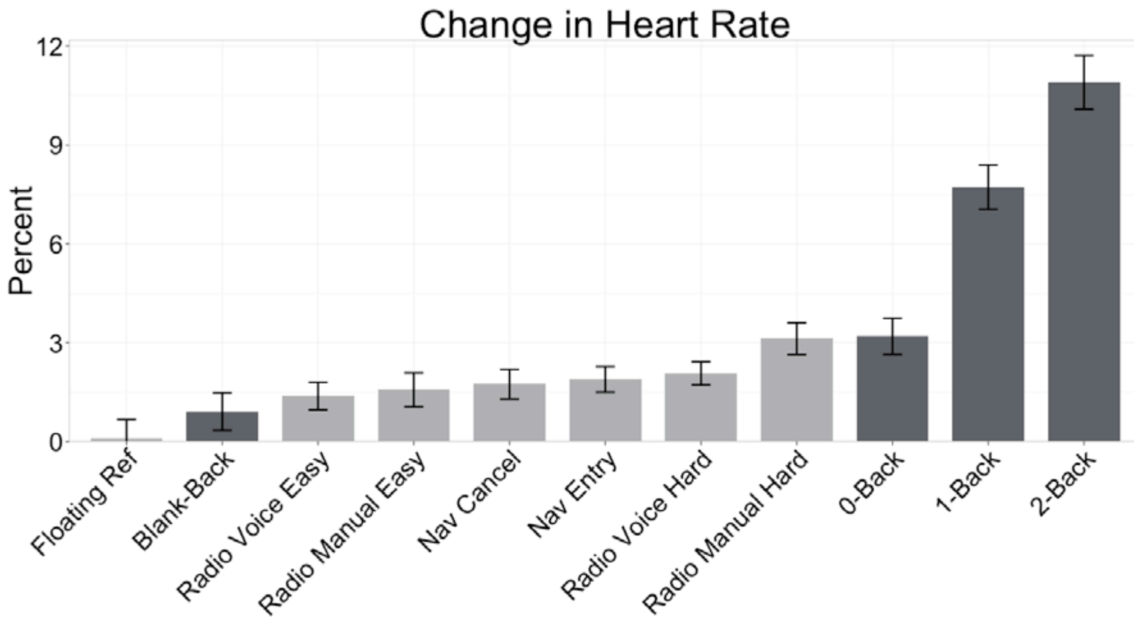


Figure 25: Change in heart rate collapsed across training conditions showing means and adjusted standard errors. Levels of the n-back task are shaded in dark gray. The "floating reference" is an interval randomly interspaced between the various n-back levels in which the participant is instructed to simply continue driving.

As in Study 1, the 1-back (M 8.3%) and 2-back (10.6%) cognitive tasks induced larger elevations in heart rate than any of the DVI tasks in the current sample. Similarly, percentage change scores across tasks are quite consistent across the two studies (i.e. for the Radio Manual Hard task - Study 1: M 3.25%, SD 5.6; Study 2:M 3.13%, SD 4.9), suggesting that although there is substantial variability in heart rate changes between subjects (figure next page), in aggregate these changes are reliable indicators of relative demand.

The Radio Manual Hard task again showed the nominally highest percentage change in heart rate of the DVI tasks, followed by Radio Voice Hard task and Nav Entry. Interestingly, while the Nav Cancel task received the highest self-reported workload rating of all DVI tasks, this does not appear in the heart rate change scores. A high physiological reaction does, however, appear in the skin conductance level increase seen during the Nav Cancel task (next measure section). This would tend to reinforce the finding that electrodermal activity (skin conductance) is more sensitive to the emotional / frustration aspects of workload.

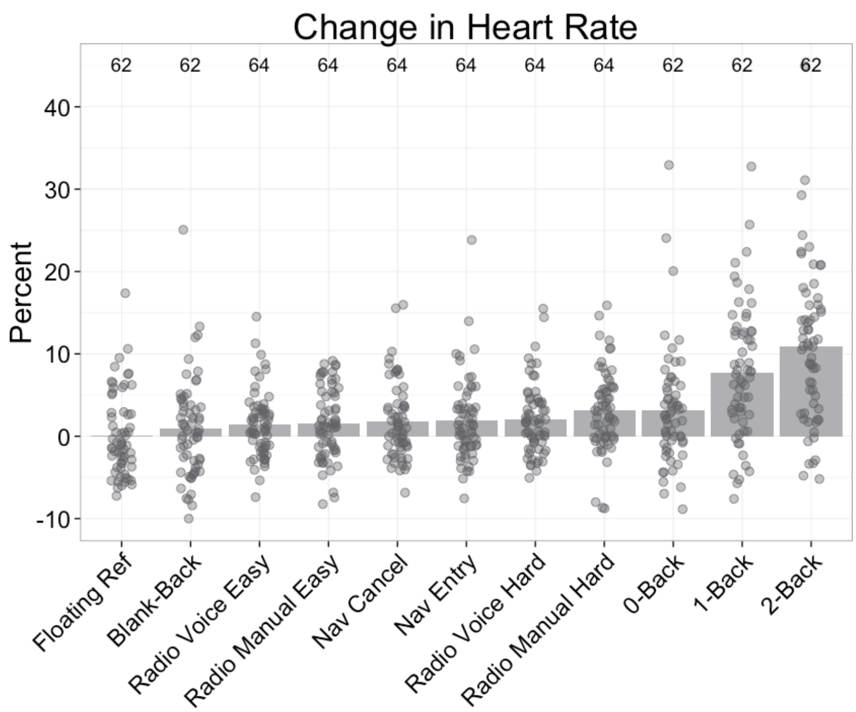


Figure 26: Percent change in heart rate with bars representing means and solid circles individual participants.

Comparison to Study 1

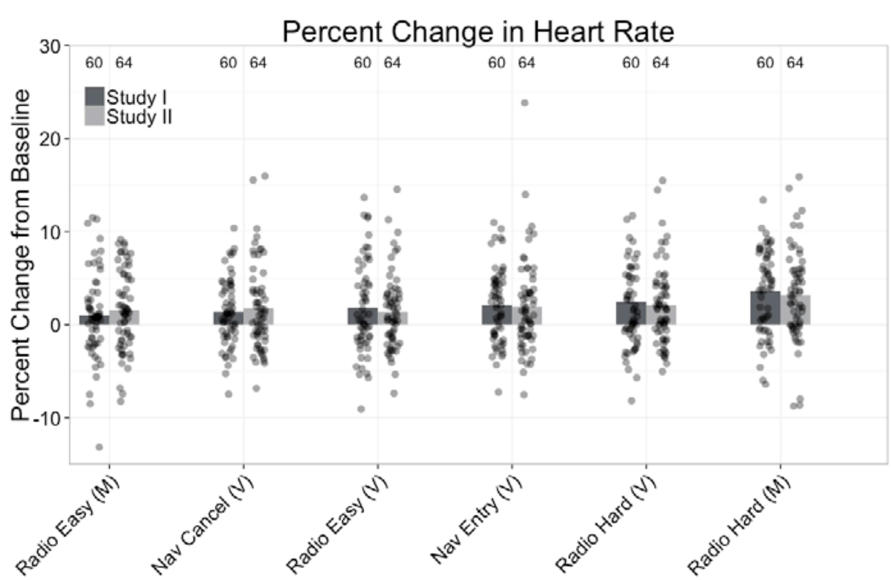


Figure 27: Comparison of percent change in heart rate in Study 1 and Study 2.

Statistical Comparison of Selected Tasks

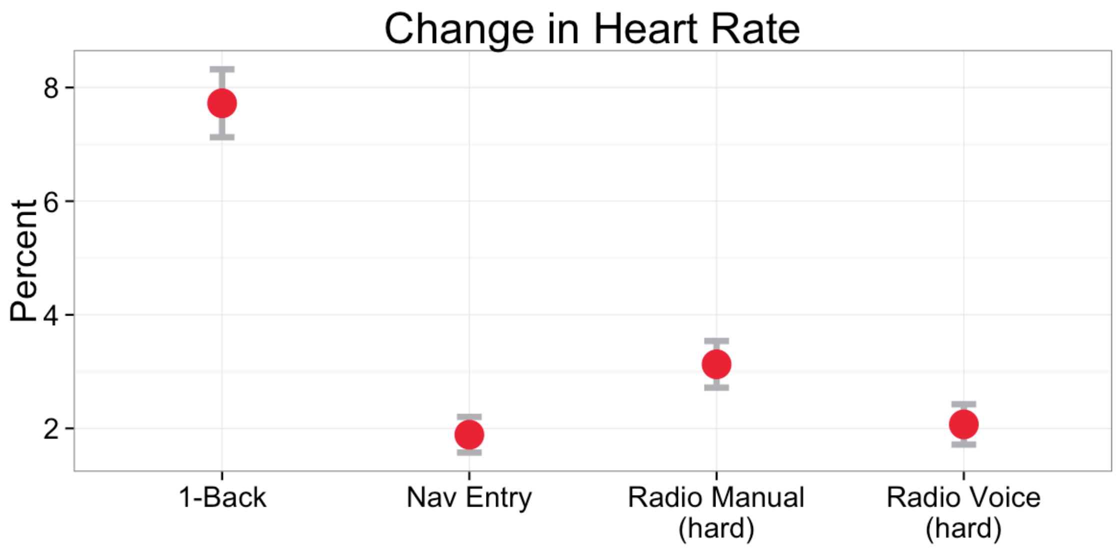


Figure 28: Percent change in heart rate across selected reference tasks.

Change in heart rate differed significantly across the four reference tasks, driven mostly by the stronger change during the 1-back task. The 1-back task induces a change in heart rate significantly greater than that of any of the other three tasks ($p < .001$ for all comparisons). Additionally, the Radio Manual Hard task was found to induce a slightly greater change in heart rate than the Navigation Entry task or the Radio Voice Hard task ($p = .034$ and $p = .080$, respectively). Compared to Study 1, the 1-back task induced a slightly larger change in heart rate (7.7% here vs. 4.5%), though since the 1-back produces the second-largest heart rate change in both reports, this pattern is broadly consistent. Also of note, the Navigation Entry task induced unremarkable net increases in heart rate. This may be related to nature of the visual demand aspects of destination entry task where the participant had, at various points, to take in visual information from the display screen when lists of city and or street names were presented for review and option selection. This may correspond to a sensory intake mode of attention management that is associated with momentary drops in heart rate activity while electrodermal activity (skin conductance level) remains high. At the same time, some of the potential load of the task may be moderated by compensatory behavior such as reducing vehicle speed as was discussed in the reports for Study 1.

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 16: Mean and (SD) of heart rate by task, broken down by age group and gender. (See Table 17 for percent change values.)

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	73.81 (10.5)	75.94 (11.1)	71.37 (10.6)	70.08 (13.8)	72.80 (11.5)
Nav Cancel	75.22 (10.1)	78.05 (10.4)	72.17 (11.2)	70.82 (14.1)	74.07 (11.6)
Radio Manual Easy	75.01 (9.6)	77.41 (11.2)	72.13 (10.6)	71.21 (14.0)	73.94 (11.5)
Radio Manual Hard	73.94 (12.1)	73.47 (10.4)	71.81 (10.1)	71.14 (14.6)	72.55 (11.7)
Radio Voice Easy	74.06 (11.5)	75.59 (11.7)	71.76 (9.5)	70.98 (13.8)	73.07 (11.6)
Radio Voice Hard	76.49 (13.9)	79.38 (11.4)	72.30 (10.5)	71.39 (14.6)	74.84 (12.8)
N-Back Reference	78.07 (12.9)	82.84 (11.9)	75.60 (11.2)	75.65 (16.4)	78.04 (13.3)
Blank-Back	81.14 (13.7)	86.28 (13.2)	78.16 (11.0)	75.66 (15.1)	80.28 (13.6)
0-Back	75.69 (10.9)	76.81 (11.8)	72.02 (10.6)	70.99 (13.5)	73.88 (11.7)
1-Back	75.53 (9.3)	77.73 (11.0)	74.04 (10.5)	72.30 (14.1)	74.90 (11.3)
2-Back	74.60 (10.9)	77.41 (10.5)	72.29 (10.3)	70.82 (15.3)	73.78 (11.9)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	72.85 (14.0)	79.16 (7.1)	75.44 (8.7)	63.83 (7.1)	72.82 (10.8)
Nav Cancel	74.63 (12.9)	80.89 (9.1)	76.13 (9.7)	65.59 (9.2)	74.31 (11.3)
Radio Manual Easy	74.47 (12.4)	80.62 (9.8)	75.47 (8.8)	66.01 (8.0)	74.14 (10.8)
Radio Manual Hard	67.99 (14.6)	77.33 (6.8)	74.95 (7.4)	64.65 (7.1)	71.45 (10.1)
Radio Voice Easy	69.53 (13.4)	79.21 (8.3)	75.71 (6.8)	65.69 (6.3)	72.74 (9.9)
Radio Voice Hard	71.48 (15.2)	81.26 (9.5)	76.98 (8.2)	65.89 (7.4)	74.06 (11.3)
N-Back Reference	72.02 (15.1)	85.41 (12.1)	80.42 (8.5)	70.03 (8.1)	77.30 (12.2)
Blank-Back	75.52 (12.0)	89.23 (11.8)	82.72 (9.1)	72.10 (8.4)	80.19 (11.9)
0-Back	75.72 (15.0)	79.83 (7.9)	76.25 (9.7)	65.80 (9.1)	74.40 (11.5)
1-Back	75.67 (12.6)	81.63 (7.4)	78.67 (9.0)	67.15 (9.8)	75.78 (10.9)
2-Back	73.85 (14.6)	80.56 (7.8)	76.28 (9.5)	64.60 (9.7)	73.82 (11.8)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	74.78 (6.0)	72.73 (13.7)	67.30 (11.3)	76.32 (16.4)	72.78 (12.3)
Nav Cancel	75.81 (7.1)	75.20 (11.5)	68.22 (11.9)	76.06 (16.7)	73.82 (12.1)
Radio Manual Easy	75.56 (6.6)	74.19 (12.3)	68.79 (11.7)	76.41 (17.1)	73.74 (12.2)
Radio Manual Hard	78.40 (8.1)	69.62 (12.3)	68.67 (11.9)	77.63 (17.6)	73.58 (13.1)
Radio Voice Easy	77.46 (9.2)	71.98 (14.0)	67.81 (10.7)	76.26 (17.5)	73.38 (13.2)
Radio Voice Hard	80.25 (12.4)	77.50 (13.5)	67.62 (11.0)	76.89 (18.2)	75.57 (14.2)
N-Back Reference	82.62 (9.7)	80.27 (12.0)	70.77 (11.9)	81.26 (20.8)	78.73 (14.4)
Blank-Back	85.36 (14.1)	83.33 (14.6)	73.61 (11.4)	79.22 (19.6)	80.38 (15.2)
0-Back	75.66 (5.2)	73.79 (14.6)	67.79 (10.1)	76.18 (15.7)	73.36 (12.0)
1-Back	75.38 (5.0)	73.83 (13.0)	69.41 (10.3)	77.44 (16.5)	74.01 (11.8)
2-Back	75.36 (6.4)	74.27 (12.3)	68.30 (10.0)	77.04 (17.8)	73.74 (12.2)

Table 17: Mean and (SD) of percent change in heart rate by task, broken down by age group and gender. (See

Combined	20-24	25-39	40-54	55+	(all)
Nav Entry	2.14 (5.1)	3.24 (6.5)	1.13 (4.5)	1.04 (3.5)	1.89 (5.0)
Nav Cancel	1.94 (5.2)	2.17 (5.3)	1.16 (3.5)	1.70 (4.7)	1.74 (4.6)
Radio Manual Easy	2.65 (4.8)	1.18 (4.3)	0.99 (4.1)	1.46 (4.2)	1.57 (4.3)
Radio Manual Hard	2.81 (6.1)	2.62 (5.2)	3.90 (4.2)	3.19 (4.2)	3.13 (4.9)
Radio Voice Easy	1.06 (4.1)	2.24 (4.0)	1.47 (4.4)	0.75 (3.9)	1.38 (4.0)
Radio Voice Hard	2.81 (4.9)	2.23 (4.2)	1.73 (4.8)	1.52 (3.4)	2.07 (4.3)
N-Back Reference	1.46 (6.8)	-3.11 (3.4)	0.80 (4.1)	1.46 (4.3)	0.11 (5.0)
Blank-Back	1.94 (9.0)	-0.54 (3.3)	0.89 (6.1)	1.48 (4.9)	0.91 (6.0)
0-Back	4.95 (11.0)	4.83 (6.6)	1.40 (4.0)	1.83 (4.7)	3.20 (7.0)
1-Back	7.30 (9.6)	9.50 (8.3)	6.21 (7.7)	7.83 (6.9)	7.72 (8.0)
2-Back	11.84 (13.6)	13.90 (8.6)	9.82 (6.8)	8.18 (7.7)	10.90 (9.4)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	2.88 (4.8)	2.04 (4.4)	0.89 (5.7)	2.50 (3.8)	2.08 (4.5)
Nav Cancel	2.77 (5.2)	1.65 (5.1)	0.02 (2.3)	3.47 (5.9)	1.98 (4.8)
Radio Manual Easy	3.92 (4.5)	0.82 (3.4)	0.99 (4.8)	2.87 (5.2)	2.15 (4.5)
Radio Manual Hard	4.63 (7.0)	3.14 (2.7)	4.38 (5.2)	4.88 (4.6)	4.26 (4.9)
Radio Voice Easy	1.26 (2.5)	1.73 (2.9)	1.07 (4.2)	0.86 (4.9)	1.23 (3.6)
Radio Voice Hard	2.02 (3.0)	1.04 (2.4)	0.60 (4.4)	3.26 (3.6)	1.73 (3.4)
N-Back Reference	-3.08 (2.2)	-2.23 (4.0)	-0.45 (4.0)	1.39 (4.7)	-0.96 (4.1)
Blank-Back	-0.43 (6.5)	0.01 (4.3)	0.81 (7.6)	3.16 (4.9)	0.97 (5.8)
0-Back	1.89 (4.8)	2.51 (5.7)	2.19 (4.1)	3.32 (5.4)	2.52 (4.8)
1-Back	3.03 (8.3)	7.55 (7.8)	6.99 (8.7)	9.74 (4.8)	7.08 (7.5)
2-Back	9.16 (13.3)	12.51 (9.0)	9.78 (5.7)	13.00 (6.0)	11.24 (8.3)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	1.40 (5.7)	4.44 (8.3)	1.37 (3.5)	-0.41 (2.6)	1.70 (5.5)
Nav Cancel	1.11 (5.5)	2.69 (5.7)	2.31 (4.3)	-0.07 (2.1)	1.51 (4.5)
Radio Manual Easy	1.37 (5.1)	1.54 (5.4)	1.00 (3.6)	0.05 (2.6)	0.99 (4.1)
Radio Manual Hard	0.99 (4.7)	2.10 (7.1)	3.42 (3.2)	1.50 (3.1)	2.00 (4.7)
Radio Voice Easy	0.86 (5.5)	2.75 (5.0)	1.87 (4.7)	0.64 (3.0)	1.53 (4.5)
Radio Voice Hard	3.59 (6.4)	3.43 (5.3)	2.86 (5.2)	-0.23 (2.4)	2.41 (5.1)
N-Back Reference	4.86 (7.2)	-3.99 (2.7)	2.06 (4.1)	1.53 (4.2)	1.11 (5.6)
Blank-Back	3.72 (10.6)	-1.09 (2.1)	0.97 (4.8)	-0.19 (4.5)	0.85 (6.3)
0-Back	7.25 (14.0)	7.15 (6.9)	0.61 (3.9)	0.34 (3.6)	3.84 (8.6)
1-Back	10.50 (9.8)	11.46 (8.8)	5.44 (7.1)	5.91 (8.5)	8.33 (8.6)
2-Back	13.85 (14.3)	15.28 (8.6)	9.86 (8.1)	3.35 (6.2)	10.58 (10.4)

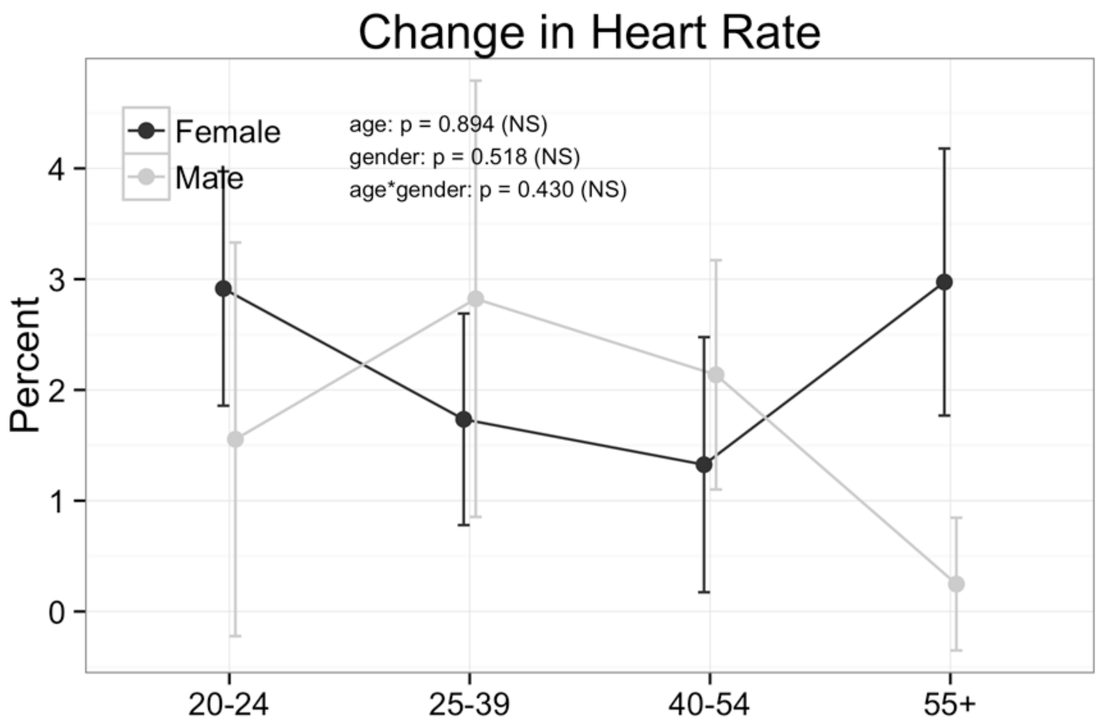


Figure 29: Percent change in heart rate across age groups and genders.

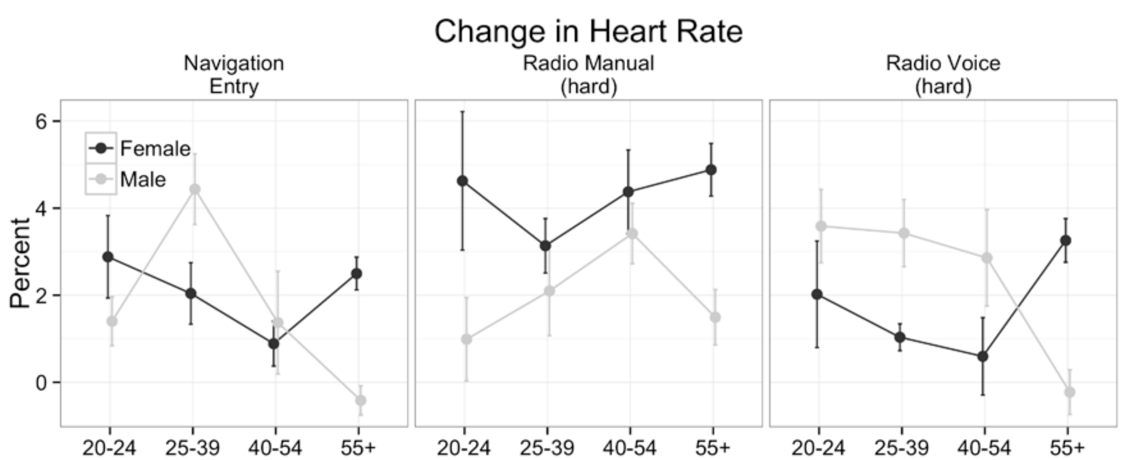


Figure 30: Percent change in heart rate by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing heart rate by age and gender, no significant main effects of age or gender appear, consistent with the lack of demographic effects for this measure in Study 1.

Skin Conductance Level (SCL)

Percent change in skin conductance level (SCL) relative to the mean of each subject's two-minute baseline periods is shown in the figures and tables below. Like changes in heart rate, changes in SCL have been shown to be sensitive to changes in cognitive demand (Mehler et al. 2012). Note that, due to excessive movement artifact in SCL recordings, data for two participants in the Structured Training condition and one in the Self-Guided Training condition could not be utilized. Movement artifact was also an issue for one participant during the n-back task periods and that person's data for those periods has similarly been discarded.

Table 18: Mean (and standard deviation) of percent change in SCL by training type.

	Structured Training	Self-Guided Training	(combined)
Nav Entry	13.60 (13.7)	8.57 (11.5)	11.04 (12.8)
Nav Cancel	15.74 (16.7)	10.12 (13.7)	12.89 (15.4)
Radio Manual Easy	6.82 (17.2)	1.75 (15.2)	4.25 (16.3)
Radio Manual Hard	12.52 (15.5)	8.33 (14.6)	10.39 (15.1)
Radio Voice Easy	7.20 (11.8)	4.63 (8.8)	5.90 (10.4)
Radio Voice Hard	5.47 (15.5)	6.94 (15.8)	6.22 (15.5)

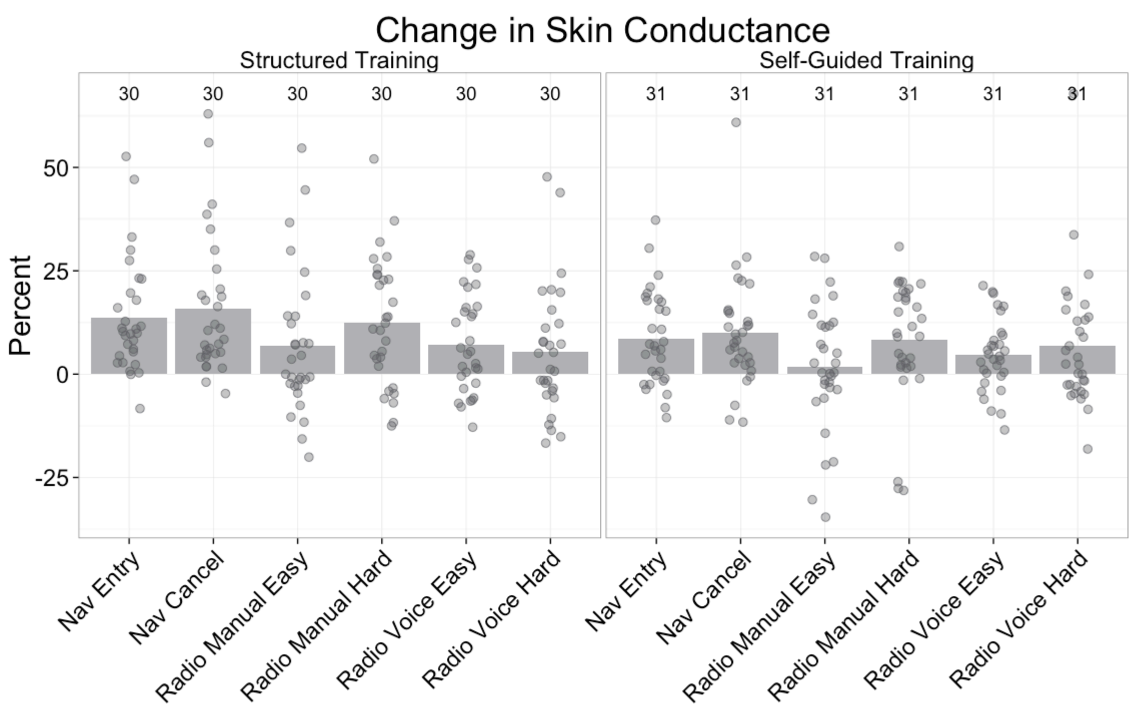


Figure 31: Percent change in SCL by task and training condition.

Change in SCL was not affected by training condition (p = .295). As in our other measures, broadly similar changes in SCL are evident in both training conditions.

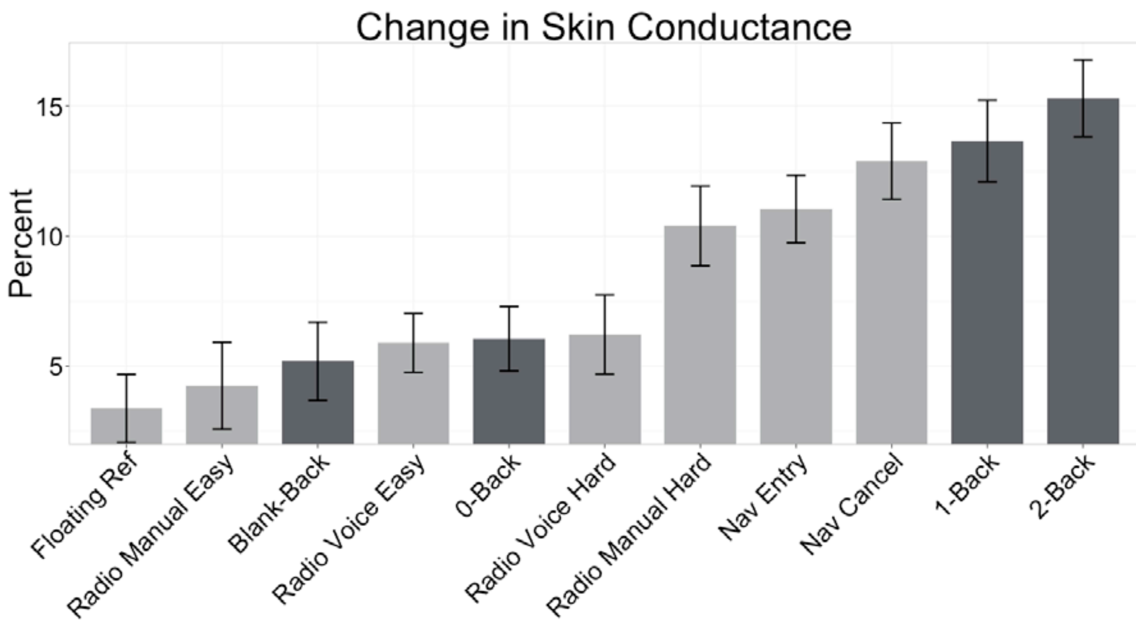


Figure 32: Percent change in skin conductance level collapsed across training conditions showing means and adjusted standard errors. Levels of the n-back task are shaded in dark gray.

As in Study 1, the arousal level, as reflected in SCL, associated with most of the DVI tasks fell between the mean values for 0-back and 1-back cognitive tasks. The Radio Voice Easy and Radio Manual Easy tasks again fell in the same nominal range as the 0-back and the Radio Manual Hard task and the Nav Entry produced SCL change scores closer to the 1-back task level. Percentage change scores across tasks are relatively consistent across the two studies, suggesting that although there is substantial variability in SCL changes between subjects, in aggregate these changes are fairly reliable indicators of relative demand.

As already discussed, the Nav Cancel task received the highest self-reported workload rating of all DVI tasks and this is concordant with SCL showing the greatest increase for all of the DVI tasks. This would tend to reinforce the position that SCL is highly sensitive to the emotional / frustration aspects of workload.

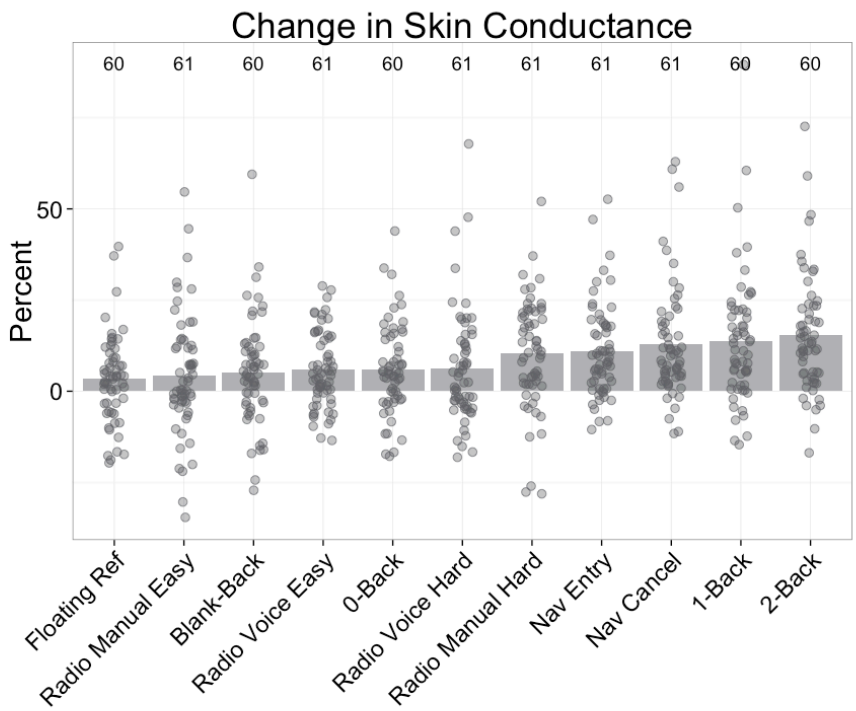


Figure 33: Percent change in skin conductance level with bars representing means and solid circles individual participants

Comparison to Study 1

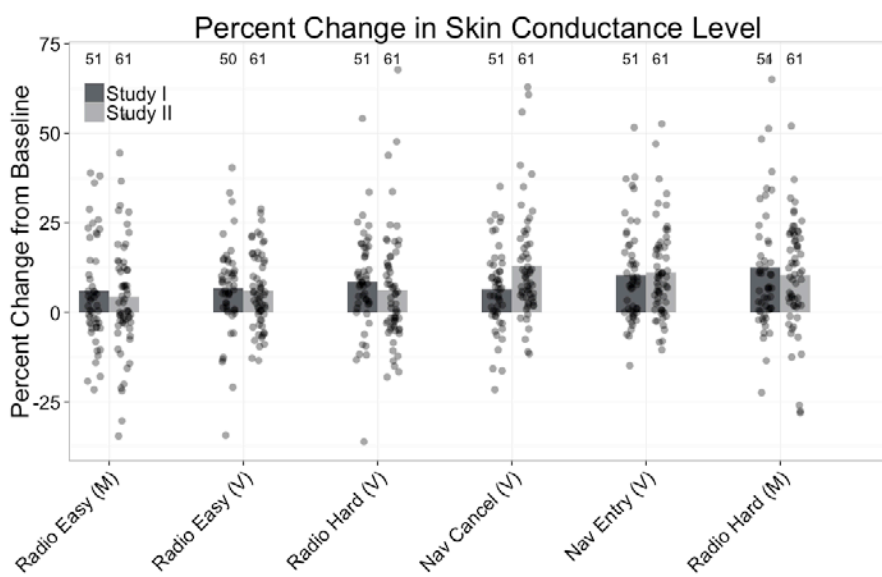


Figure 34: Comparison of percent change in SCL in Study 1 and Study 2.

It can be observed in Figure 34 that the relative increase in SCL during the Nav Cancel task was markedly higher in Study 2 than in Study 1. This goes along with the previously discussed observation that many participants found this task harder in

Study 2 as the task prompt did not explicitly remind the participant of the specific verbal command structure to use to cancel the task.

Statistical Comparison of Selected Tasks

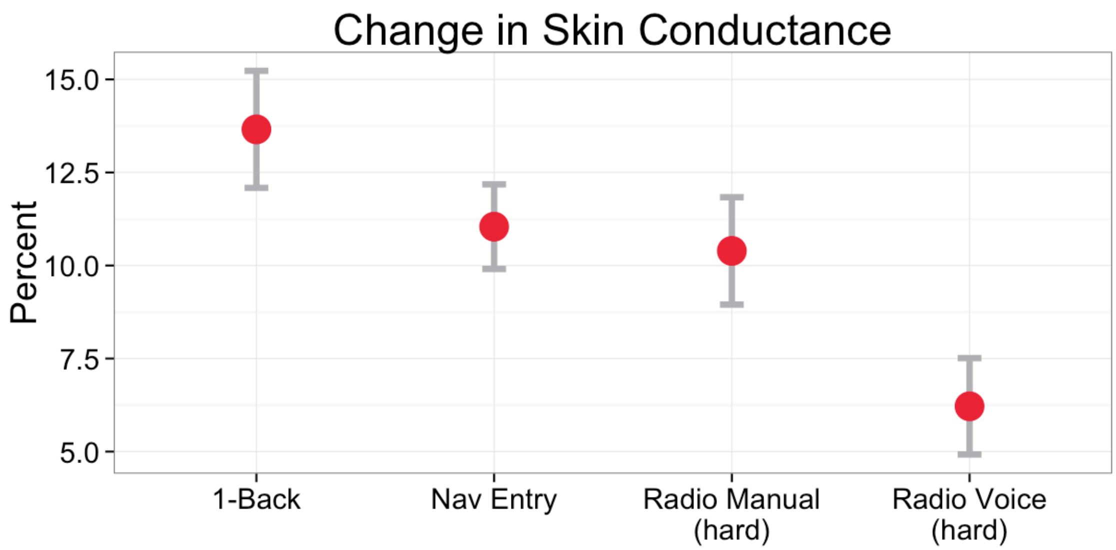


Figure 35: Percentage change in SCL across selected reference tasks.

Change in SCL differed significantly across the four reference tasks. The 1-back, Navigation Entry, and Radio Manual Hard tasks all show comparable increases in SCL compared to baseline, whereas SCL change during the Radio Voice Hard task is relatively small. SCL changes during the Radio Voice task are significantly smaller compared to all other tasks ($p < .01$ for all comparisons, Wilcoxon test). There were no other significant differences between reference tasks. This is consistent with the results of Study 1, which similarly found that Radio Voice SCL was significantly lower than the other reference tasks.

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 19: Mean and (SD) of SCL (in micromhos) by task, broken down by age group and gender. (See Table 20 for percent change values.)

Combined	20-24	25-39	40-54	55+	(all)
Nav Entry	11.98 (5.3)	10.00 (3.4)	8.41 (3.5)	6.71 (2.7)	9.29 (4.2)
Nav Cancel	13.38 (7.4)	11.04 (3.5)	9.38 (3.5)	7.00 (2.5)	10.15 (5.0)
Radio Manual Easy	13.94 (9.5)	11.16 (3.5)	9.53 (3.4)	7.07 (2.6)	10.37 (5.8)
Radio Manual Hard	11.47 (4.6)	10.61 (3.7)	8.50 (4.0)	7.25 (3.2)	9.44 (4.1)
Radio Voice Easy	11.45 (4.1)	10.60 (4.2)	8.75 (3.7)	7.33 (2.6)	9.52 (3.9)
Radio Voice Hard	11.99 (5.1)	10.86 (3.9)	8.72 (3.8)	7.31 (3.2)	9.70 (4.3)
N-Back Reference	12.57 (6.2)	11.58 (4.4)	9.44 (4.4)	7.93 (2.9)	10.36 (4.8)
Blank-Back	12.68 (5.7)	11.34 (4.0)	9.69 (4.0)	7.95 (2.5)	10.39 (4.4)
0-Back	12.42 (6.6)	10.46 (3.8)	8.90 (3.7)	6.75 (2.7)	9.59 (4.8)
1-Back	12.84 (6.4)	11.05 (4.2)	9.61 (3.8)	7.04 (2.5)	10.09 (4.8)
2-Back	12.54 (6.4)	10.40 (3.5)	9.21 (3.7)	6.76 (2.5)	9.69 (4.6)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	10.49 (2.8)	8.65 (3.4)	7.02 (2.6)	7.18 (3.9)	8.35 (3.3)
Nav Cancel	11.39 (3.4)	9.66 (3.3)	7.67 (2.6)	7.09 (3.5)	8.87 (3.5)
Radio Manual Easy	11.51 (3.4)	9.85 (3.5)	7.83 (2.4)	7.22 (3.6)	9.02 (3.5)
Radio Manual Hard	9.30 (1.5)	9.25 (3.7)	6.86 (2.9)	7.52 (4.4)	8.23 (3.3)
Radio Voice Easy	9.43 (1.3)	9.01 (3.9)	7.47 (2.7)	7.69 (3.4)	8.38 (3.0)
Radio Voice Hard	9.63 (1.5)	9.41 (4.0)	7.37 (3.0)	7.79 (4.4)	8.54 (3.5)
N-Back Reference	10.01 (2.4)	10.03 (3.7)	7.50 (3.2)	8.04 (3.7)	8.90 (3.4)
Blank-Back	10.05 (1.9)	9.69 (3.8)	7.96 (2.8)	7.92 (3.2)	8.89 (3.1)
0-Back	10.92 (2.8)	8.76 (3.6)	7.53 (2.8)	6.95 (3.6)	8.46 (3.4)
1-Back	11.13 (2.7)	9.60 (3.9)	8.11 (3.0)	7.04 (3.5)	8.90 (3.5)
2-Back	10.69 (2.8)	9.13 (3.4)	7.29 (2.6)	6.93 (3.5)	8.48 (3.3)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	13.29 (6.8)	11.35 (3.0)	9.63 (3.9)	6.30 (1.3)	10.14 (4.8)
Nav Cancel	15.13 (9.5)	12.41 (3.3)	11.10 (3.6)	6.92 (1.0)	11.39 (5.9)
Radio Manual Easy	16.07 (12.6)	12.47 (3.2)	11.23 (3.5)	6.92 (1.2)	11.67 (7.2)
Radio Manual Hard	13.09 (5.6)	11.98 (3.4)	9.93 (4.4)	7.02 (2.1)	10.51 (4.5)
Radio Voice Easy	12.97 (4.9)	12.20 (4.1)	9.88 (4.2)	7.01 (1.7)	10.51 (4.4)
Radio Voice Hard	13.76 (6.2)	12.31 (3.4)	9.90 (4.2)	6.89 (1.9)	10.72 (4.8)
N-Back Reference	14.49 (7.5)	13.12 (4.6)	11.14 (4.8)	7.83 (2.4)	11.64 (5.5)
Blank-Back	14.66 (6.8)	12.99 (3.7)	11.22 (4.4)	7.97 (1.9)	11.71 (5.0)
0-Back	13.73 (8.7)	12.15 (3.5)	10.27 (4.0)	6.55 (1.6)	10.67 (5.6)
1-Back	14.34 (8.4)	12.50 (4.1)	11.11 (4.1)	7.04 (1.4)	11.25 (5.6)
2-Back	14.16 (8.3)	11.67 (3.3)	10.88 (3.8)	6.58 (0.9)	10.82 (5.4)

Table 20: Mean and (SD) of percent change in SCL by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	9.42 (8.7)	12.18 (11.8)	12.79 (15.1)	9.70 (15.4)	11.04 (12.8)
Nav Cancel	11.57 (17.3)	13.21 (14.5)	15.82 (17.8)	10.93 (12.6)	12.89 (15.4)
Radio Manual Easy	2.81 (12.2)	4.13 (16.6)	4.88 (16.6)	5.18 (20.4)	4.25 (16.3)
Radio Manual Hard	7.12 (13.1)	9.94 (17.1)	13.74 (14.0)	10.79 (16.3)	10.39 (15.1)
Radio Voice Easy	3.42 (9.2)	4.65 (6.8)	10.71 (13.0)	4.89 (11.2)	5.90 (10.4)
Radio Voice Hard	5.97 (19.1)	8.31 (14.1)	6.30 (11.3)	4.14 (17.9)	6.22 (15.5)
N-Back Reference	-0.40 (8.8)	6.08 (7.1)	-0.13 (14.2)	7.52 (14.7)	3.38 (11.9)
Blank-Back	0.04 (8.3)	3.83 (10.7)	4.52 (14.9)	12.08 (19.9)	5.18 (14.5)
0-Back	3.38 (11.3)	8.33 (10.8)	3.46 (12.2)	8.73 (15.2)	6.06 (12.4)
1-Back	6.55 (10.1)	16.08 (15.3)	10.92 (15.6)	20.45 (24.6)	13.66 (17.6)
2-Back	8.83 (12.6)	13.16 (7.1)	16.30 (16.9)	22.62 (22.7)	15.30 (16.2)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	7.77 (8.4)	14.86 (14.7)	5.71 (7.8)	7.95 (20.2)	9.27 (13.5)
Nav Cancel	9.12 (12.7)	16.18 (18.1)	9.64 (9.8)	11.02 (14.7)	11.65 (13.8)
Radio Manual Easy	4.35 (9.1)	1.20 (21.5)	1.12 (7.7)	5.62 (22.7)	3.01 (16.2)
Radio Manual Hard	6.74 (9.7)	10.92 (22.8)	9.69 (11.4)	7.51 (13.5)	8.79 (14.8)
Radio Voice Easy	2.28 (12.1)	6.72 (8.5)	5.79 (14.6)	3.52 (11.4)	4.65 (11.2)
Radio Voice Hard	-0.97 (9.0)	12.17 (18.4)	1.81 (7.3)	6.04 (23.3)	5.02 (16.0)
N-Back Reference	-2.55 (9.9)	6.83 (8.0)	-2.96 (10.6)	3.64 (7.5)	1.57 (9.5)
Blank-Back	-1.40 (5.2)	1.53 (11.1)	6.74 (10.4)	11.20 (12.1)	4.62 (10.8)
0-Back	1.24 (14.6)	8.30 (15.2)	4.11 (10.9)	8.14 (9.8)	5.70 (12.5)
1-Back	3.92 (13.7)	18.22 (18.7)	6.00 (10.6)	15.32 (13.8)	11.37 (15.2)
2-Back	5.40 (17.1)	12.17 (7.8)	14.37 (7.0)	17.33 (20.4)	12.56 (13.8)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	10.86 (9.3)	9.50 (8.3)	18.99 (17.5)	11.23 (11.0)	12.65 (12.1)
Nav Cancel	13.72 (21.2)	10.25 (10.3)	21.24 (21.8)	10.84 (11.5)	14.01 (16.8)
Radio Manual Easy	1.46 (15.0)	7.05 (10.4)	8.17 (21.7)	4.79 (19.7)	5.37 (16.6)
Radio Manual Hard	7.46 (16.2)	8.96 (10.4)	17.29 (15.7)	13.66 (18.9)	11.84 (15.4)
Radio Voice Easy	4.41 (6.3)	2.57 (4.2)	15.01 (10.5)	6.10 (11.6)	7.02 (9.6)
Radio Voice Hard	12.04 (23.8)	4.45 (7.3)	10.23 (13.1)	2.48 (13.0)	7.30 (15.3)
N-Back Reference	1.21 (8.2)	5.34 (6.6)	2.35 (17.0)	10.92 (18.8)	4.95 (13.6)
Blank-Back	1.12 (10.2)	6.12 (10.5)	2.58 (18.5)	12.86 (25.8)	5.67 (17.2)
0-Back	5.00 (8.8)	8.36 (4.5)	2.89 (13.9)	9.26 (19.5)	6.38 (12.6)
1-Back	8.52 (6.8)	13.94 (11.8)	15.23 (18.6)	24.94 (31.6)	15.66 (19.5)
2-Back	11.40 (8.3)	14.16 (6.7)	17.99 (22.8)	27.24 (25.0)	17.70 (17.9)

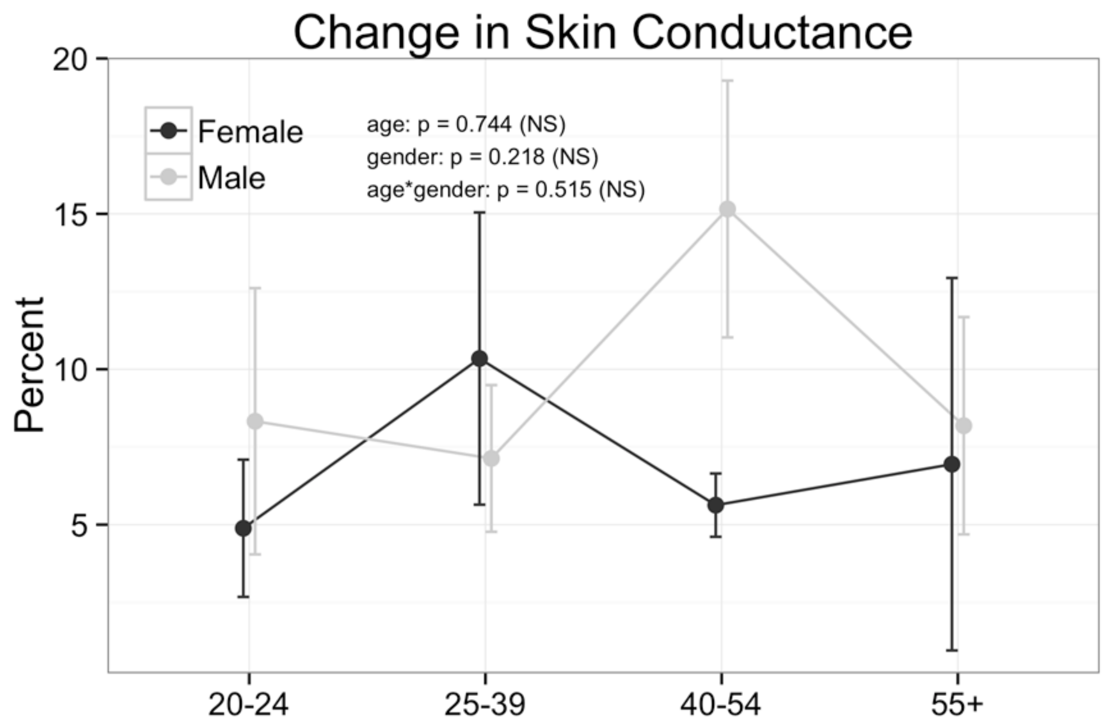


Figure 36: Percent change in SCL across age groups and genders.

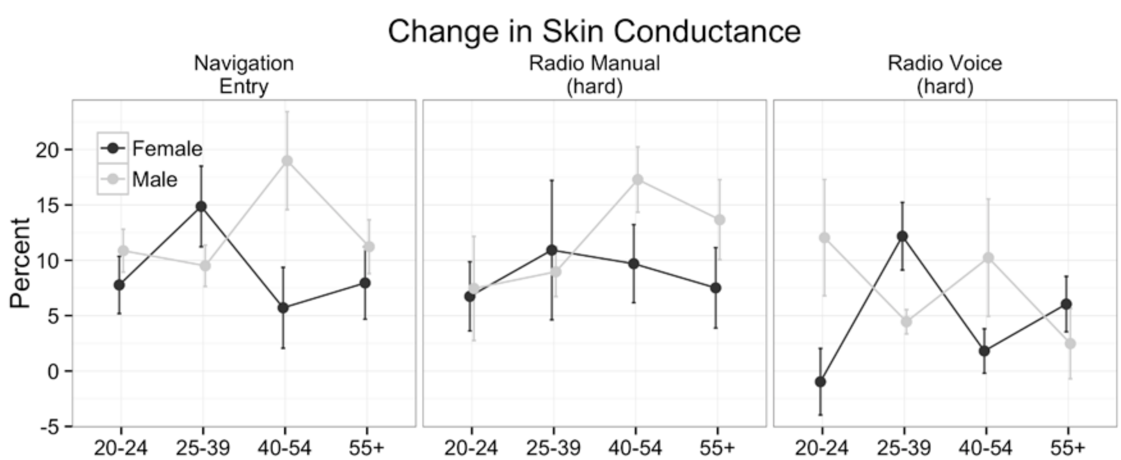


Figure 37: Percent change in SCL by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing SCL by age and gender, no statistically significant age or gender effects appear.

Driving Behavior Measures

Mean Velocity

Reductions in speed relative to single task driving are often interpreted as compensatory behaviors to reduce workload and/or increase safety margins (Angell, et al., 2006; Horberry, Anderson, Regan, Triggs, & Brown, 2006; Lerner, Singer, & Huey, 2008).

Table 21: Mean (and standard deviation) of mean vehicle velocity by training type.

	Structured Training	Self-Guided Training	(combined)
Baseline	107.85 (7.5)	108.23 (6.8)	108.04 (7.1)
Nav Entry	108.01 (6.7)	107.72 (9.4)	107.87 (8.1)
Nav Cancel	107.63 (8.1)	106.68 (12.3)	107.16 (10.3)
Radio Manual Easy	107.17 (9.2)	108.12 (10.2)	107.65 (9.6)
Radio Manual Hard	106.20 (7.4)	105.05 (8.7)	105.63 (8.0)
Radio Voice Easy	107.44 (7.7)	107.37 (11.2)	107.41 (9.5)
Radio Voice Hard	107.96 (8.4)	108.49 (9.8)	108.23 (9.0)

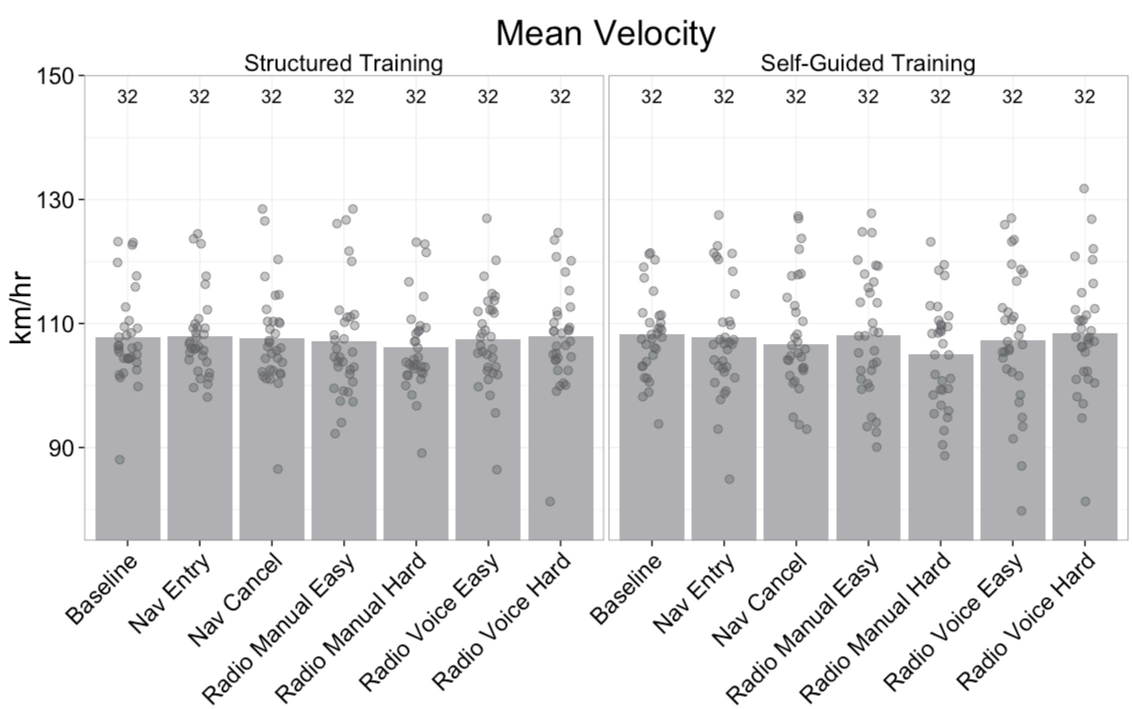


Figure 38: Mean vehicle velocity by task and training condition.

Mean vehicle velocity was not affected by training condition (p = .903).

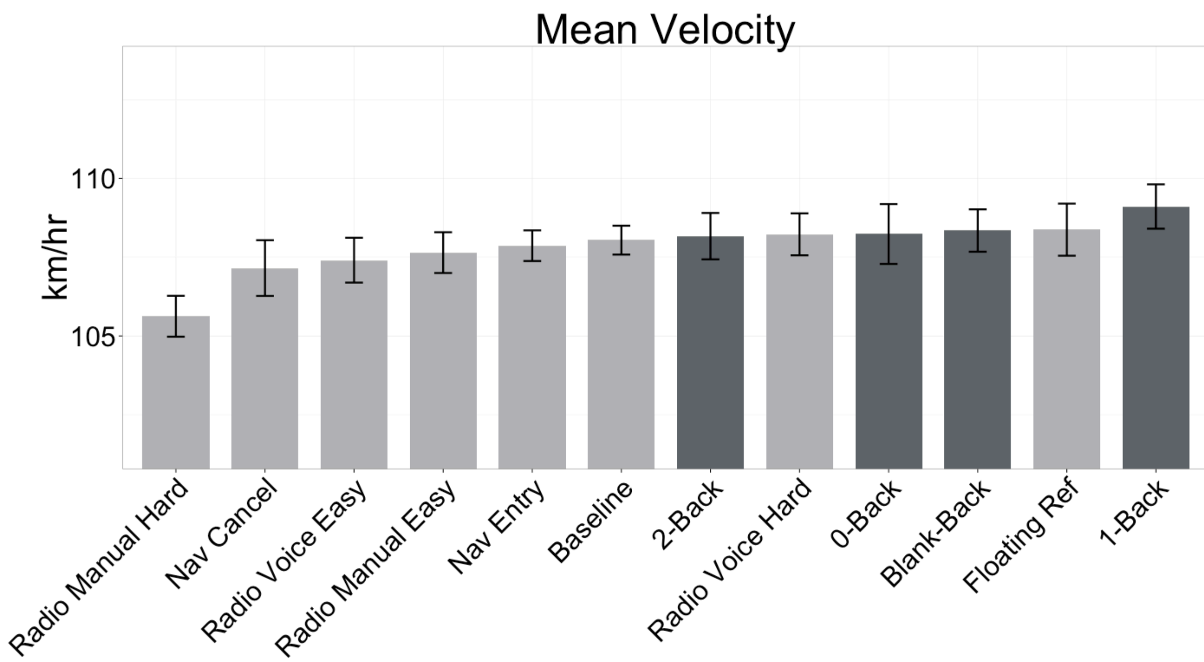


Figure 39: Mean velocity collapsed across training conditions showing means and adjusted standard errors. Levels of the n-back task are shaded in dark gray.

The variation in mean driving speed across tasks was not as pronounced in the current sample as was observed in Study 1 (105-109 vs. 105-113 km/hr respectively). Nominal vehicle speed was again highest during the low to moderately cognitively demanding n-back tasks (1-back and 0-back, and along with the new blank back task). As in Study 1, the Radio Manual Hard tuning task, which is the most manually intensive of all the DVI tasks, is associated with marked slowing relative to other DVI tasks.

As described in the *Methods* section, the "Floating Ref" is a task period of equal duration as the n-back periods that is presented in random order amongst the four levels of the n-back tasks. During this period the participant is instructed to just continue driving. This allows for consideration of the potential impact of having the driver being presented with a recorded "task instruction" even if the instruction requests nothing beyond continuing to drive. As can be observed in Figure 39 above, this floating single task reference period falls within the same velocity range as the formal baseline period.

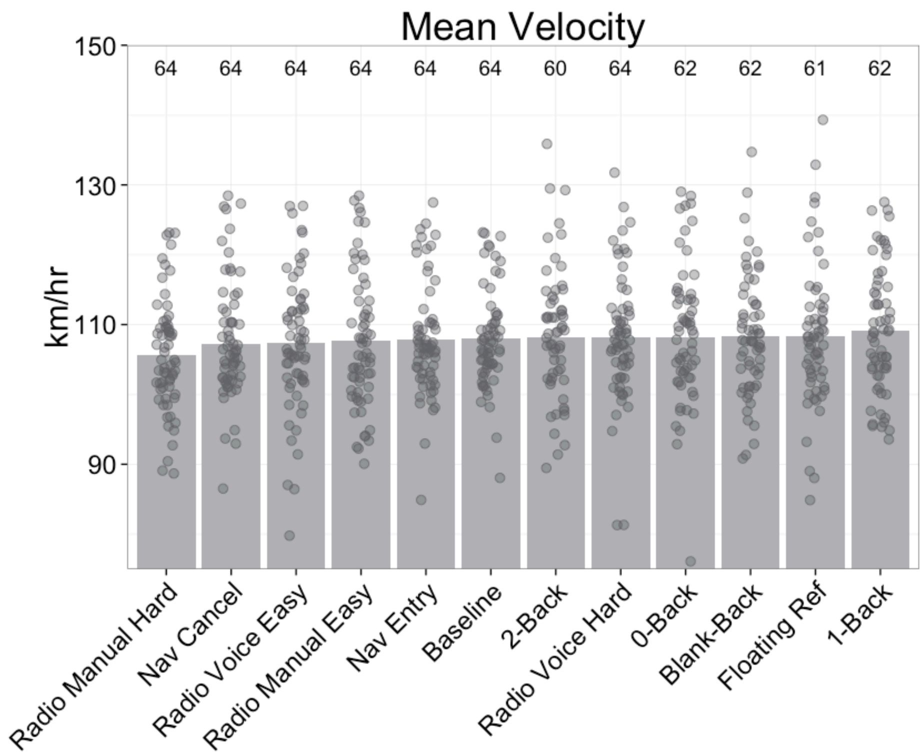


Figure 40: Mean vehicle velocity with bars representing means and solid circles individual participants.

Comparison to Study 1

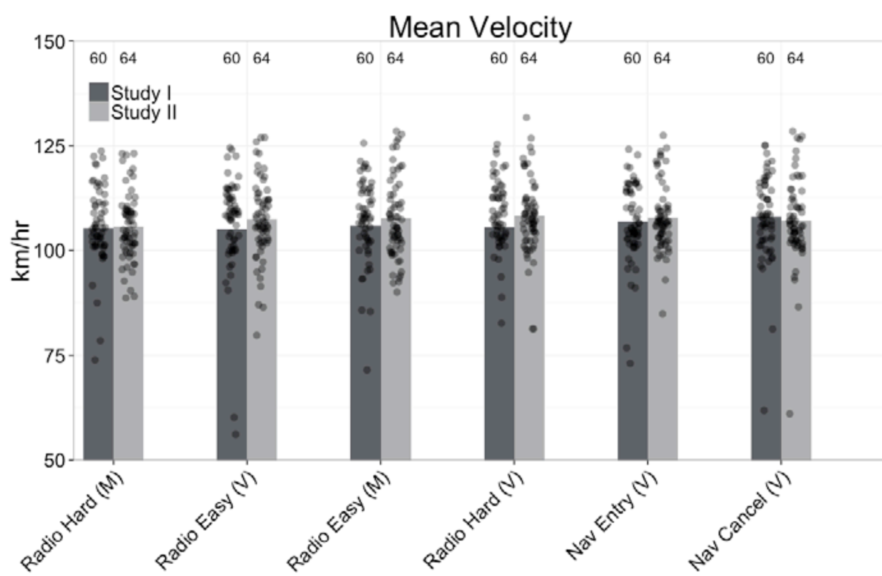


Figure 41: Comparison of mean velocity in Study 1 and Study 2.

Statistical Comparison of Selected Tasks

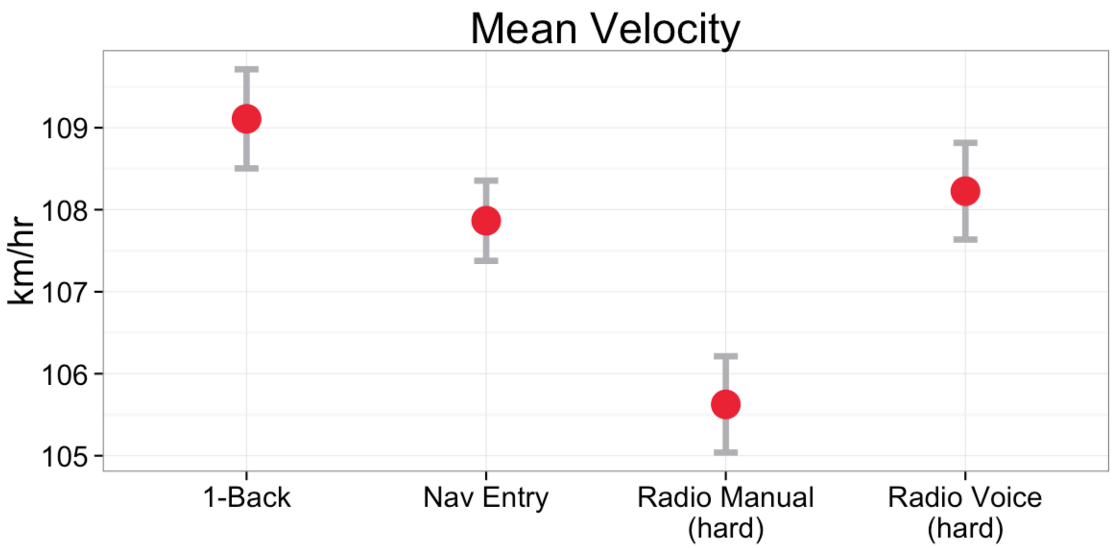


Figure 42: Mean vehicle velocity across selected reference tasks.

Mean vehicle speed varied significantly across the four reference tasks ($X^2 = 13.5$, $p = .004$, Friedman test). This effect was driven entirely by the Radio Manual Hard task, which had a significantly reduced speed compared with the 1-back, Navigation Entry, and Radio Voice tasks ($p < .001$, $p = .013$, and $p = .004$, respectively).

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 22: Mean and (SD) of vehicle velocity by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	110.47 (7.2)	110.94 (7.6)	106.56 (5.8)	104.19 (5.9)	108.04 (7.1)
Nav Entry	111.04 (7.9)	110.10 (10.0)	106.33 (7.4)	103.98 (4.8)	107.87 (8.1)
Nav Cancel	110.85 (7.9)	110.67 (11.3)	106.23 (5.3)	100.87 (12.6)	107.16 (10.3)
Radio Manual Easy	110.53 (9.1)	109.99 (12.1)	107.23 (6.9)	102.83 (8.5)	107.65 (9.6)
Radio Manual Hard	108.23 (8.8)	106.94 (8.2)	104.35 (9.2)	102.98 (4.8)	105.63 (8.0)
Radio Voice Easy	112.75 (8.0)	109.31 (11.0)	106.41 (7.8)	101.15 (7.6)	107.41 (9.5)
Radio Voice Hard	111.51 (6.2)	109.63 (13.1)	107.18 (8.4)	104.58 (5.6)	108.23 (9.0)
N-Back Reference	111.52 (12.7)	111.37 (10.9)	106.73 (7.8)	104.45 (7.5)	108.37 (10.1)
Blank-Back	110.31 (7.1)	110.81 (11.1)	109.17 (5.9)	103.34 (8.2)	108.35 (8.7)
0-Back	111.40 (8.7)	107.62 (13.7)	109.51 (8.6)	104.81 (11.7)	108.24 (11.0)
1-Back	111.54 (8.9)	110.05 (9.3)	110.56 (9.6)	104.58 (7.1)	109.11 (9.0)
2-Back	108.19 (8.0)	113.36 (12.2)	108.60 (7.0)	102.85 (11.2)	108.17 (10.3)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	107.60 (7.1)	109.81 (5.5)	107.24 (7.5)	104.74 (3.9)	107.35 (6.2)
Nav Entry	108.09 (7.1)	109.05 (7.0)	105.20 (10.1)	103.02 (4.5)	106.34 (7.5)
Nav Cancel	108.11 (8.9)	109.14 (9.4)	106.34 (7.0)	98.22 (15.1)	105.46 (10.9)
Radio Manual Easy	107.53 (8.3)	106.49 (10.4)	108.88 (7.3)	102.86 (8.3)	106.44 (8.5)
Radio Manual Hard	104.87 (7.6)	103.66 (3.7)	103.65 (9.8)	102.71 (5.7)	103.72 (6.8)
Radio Voice Easy	110.53 (8.2)	110.17 (5.6)	109.19 (7.1)	101.24 (9.8)	107.78 (8.4)
Radio Voice Hard	109.96 (5.4)	107.78 (10.3)	110.62 (5.0)	103.91 (6.8)	108.07 (7.3)
N-Back Reference	107.27 (2.6)	107.22 (7.3)	107.51 (9.1)	105.29 (4.7)	106.78 (6.3)
Blank-Back	106.89 (2.9)	106.59 (6.0)	107.95 (6.8)	102.96 (7.8)	106.04 (6.3)
0-Back	105.94 (5.3)	107.65 (16.4)	106.87 (9.8)	105.12 (7.0)	106.43 (10.3)
1-Back	107.52 (2.0)	109.45 (7.5)	112.90 (10.4)	103.49 (6.9)	108.40 (8.0)
2-Back	103.54 (6.8)	113.78 (11.6)	108.58 (8.8)	102.68 (9.5)	107.29 (10.0)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	113.34 (6.6)	112.06 (9.4)	105.89 (3.7)	103.65 (7.6)	108.73 (7.9)
Nav Entry	113.99 (7.9)	111.16 (12.8)	107.47 (3.5)	104.94 (5.1)	109.39 (8.5)
Nav Cancel	113.58 (6.0)	112.20 (13.4)	106.11 (3.5)	103.53 (9.8)	108.85 (9.5)
Radio Manual Easy	113.54 (9.3)	113.50 (13.4)	105.58 (6.5)	102.79 (9.2)	108.85 (10.6)
Radio Manual Hard	111.60 (9.1)	110.21 (10.3)	105.05 (9.1)	103.25 (4.1)	107.53 (8.8)
Radio Voice Easy	114.96 (7.8)	108.45 (15.0)	103.64 (8.0)	101.06 (5.1)	107.03 (10.7)
Radio Voice Hard	113.06 (7.0)	111.49 (16.0)	103.75 (9.9)	105.24 (4.4)	108.38 (10.6)
N-Back Reference	114.70 (16.4)	115.01 (12.6)	105.95 (6.8)	103.62 (9.8)	109.82 (12.5)
Blank-Back	112.87 (8.4)	115.03 (13.7)	110.38 (5.0)	103.73 (9.1)	110.50 (10.0)
0-Back	115.49 (8.8)	107.59 (11.5)	112.15 (6.8)	104.49 (15.6)	109.93 (11.5)
1-Back	114.55 (10.9)	110.66 (11.3)	108.22 (8.7)	105.67 (7.6)	109.78 (9.9)
2-Back	111.10 (7.7)	112.99 (13.5)	108.63 (5.3)	103.01 (13.3)	108.93 (10.7)

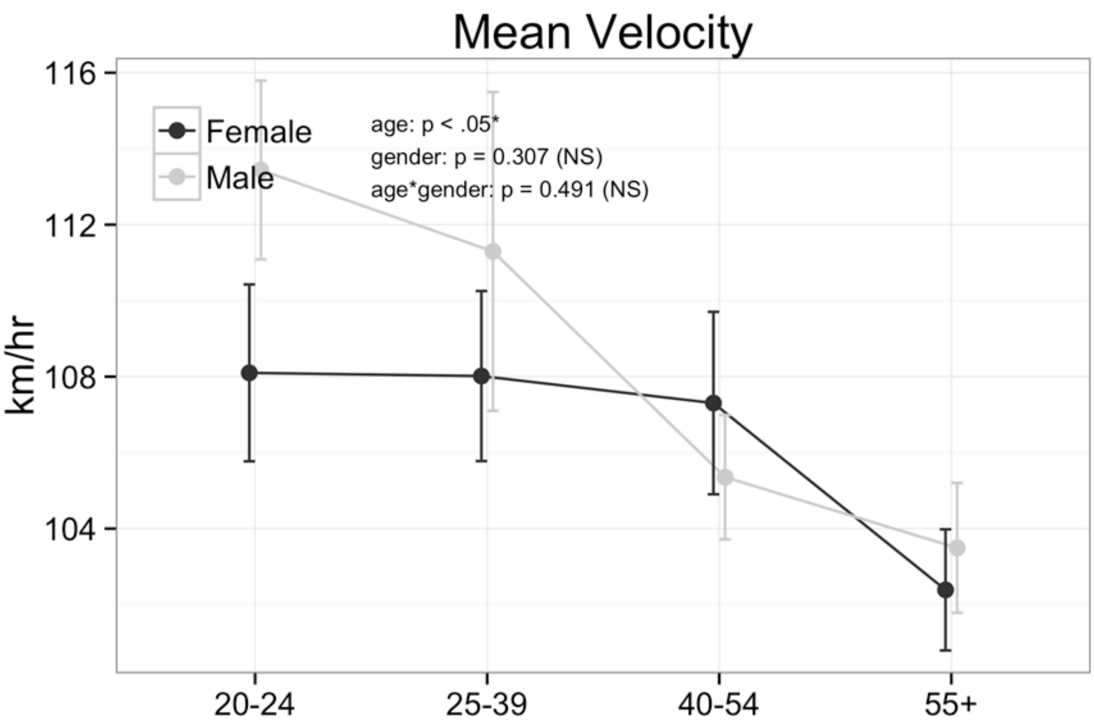


Figure 43: Mean vehicle velocity across age groups and genders.

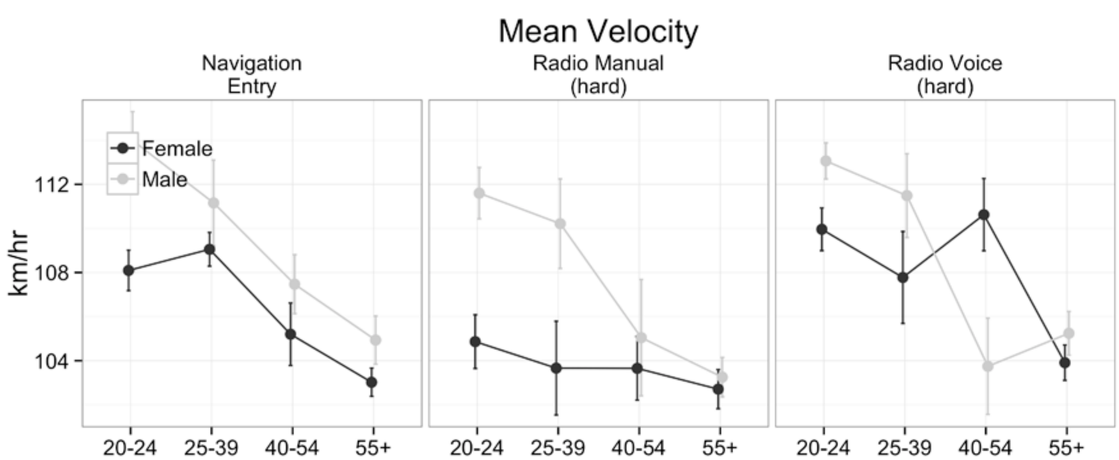


Figure 44: Mean vehicle velocity by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

The figures above indicate a significant effect of age on mean vehicle velocity, with vehicle speed decreasing with age. The effect is especially pronounced in men aged 40 and older. Although the average driving speed of female drivers also decreases with age, their driving speeds are somewhat slower (and more consistent) overall.

Variability of Velocity (Standard Deviation)

Like mean vehicle velocity, the variability of vehicle velocity (calculated as the standard deviation of vehicle speed) can reflect difference in task demands. When interpreting these data, it is important to keep in mind that the standard deviation will be affected to some extent by the amount of data (task duration) considered in the calculation. In other words, the length of the task can, in certain cases, affect the measure of variability regardless of any meaningful changes in driver control.

Table 23: Mean (and standard deviation) variability of velocity by training type.

	Structured Training	Self-Guided Training	(combined)
Baseline	4.77 (1.8)	4.39 (1.1)	4.58 (1.5)
Nav Entry	3.93 (1.5)	3.87 (1.3)	3.90 (1.4)
Nav Cancel	3.20 (1.7)	2.98 (1.5)	3.09 (1.6)
Radio Manual Easy	0.93 (0.7)	0.84 (0.4)	0.89 (0.6)
Radio Manual Hard	2.58 (1.6)	2.62 (1.1)	2.60 (1.3)
Radio Voice Easy	2.06 (1.0)	2.42 (1.6)	2.24 (1.4)
Radio Voice Hard	2.53 (1.3)	2.63 (1.2)	2.58 (1.3)

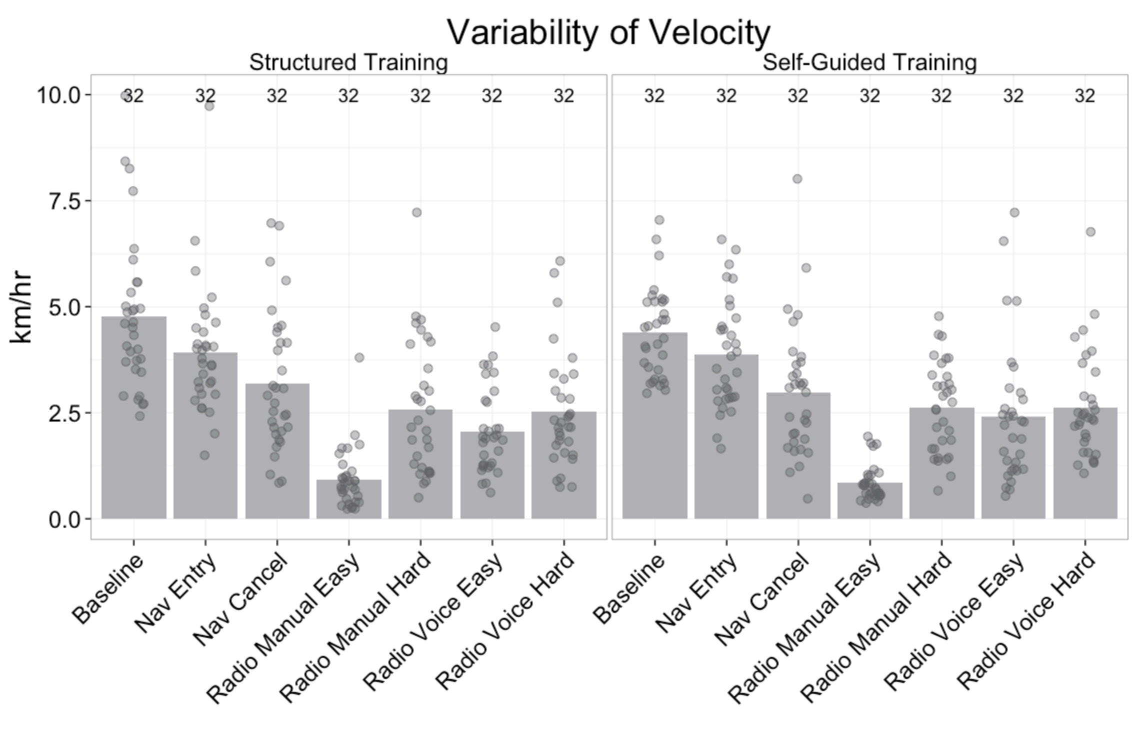


Figure 45: Variability (SD) of velocity by task and training condition.

Variability of velocity was not affected by training condition (p = .855). As with many of the other measures in this report, the consistent pattern of the variability measure between training conditions is noteworthy.

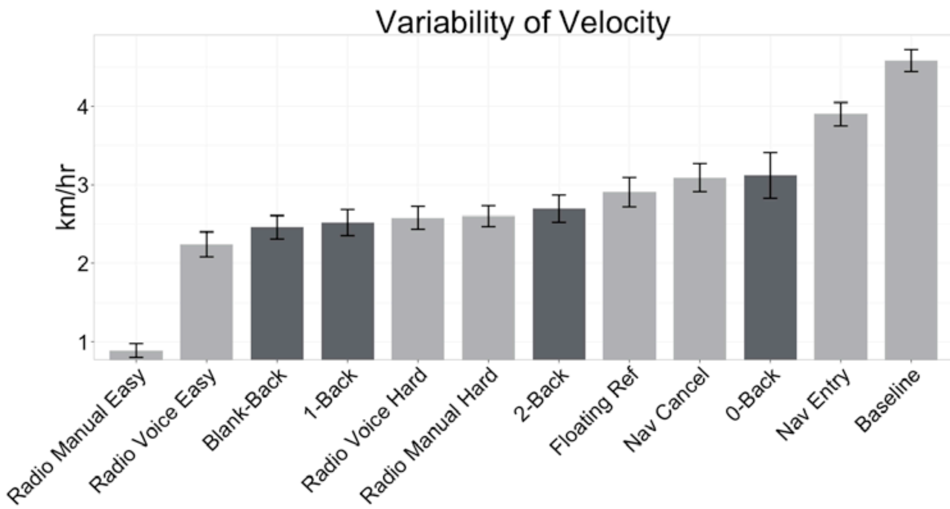


Figure 46: Variability (SD) of velocity normalized for task duration showing means and adjusted standard errors. Levels of the n-back task are shaded in dark gray.

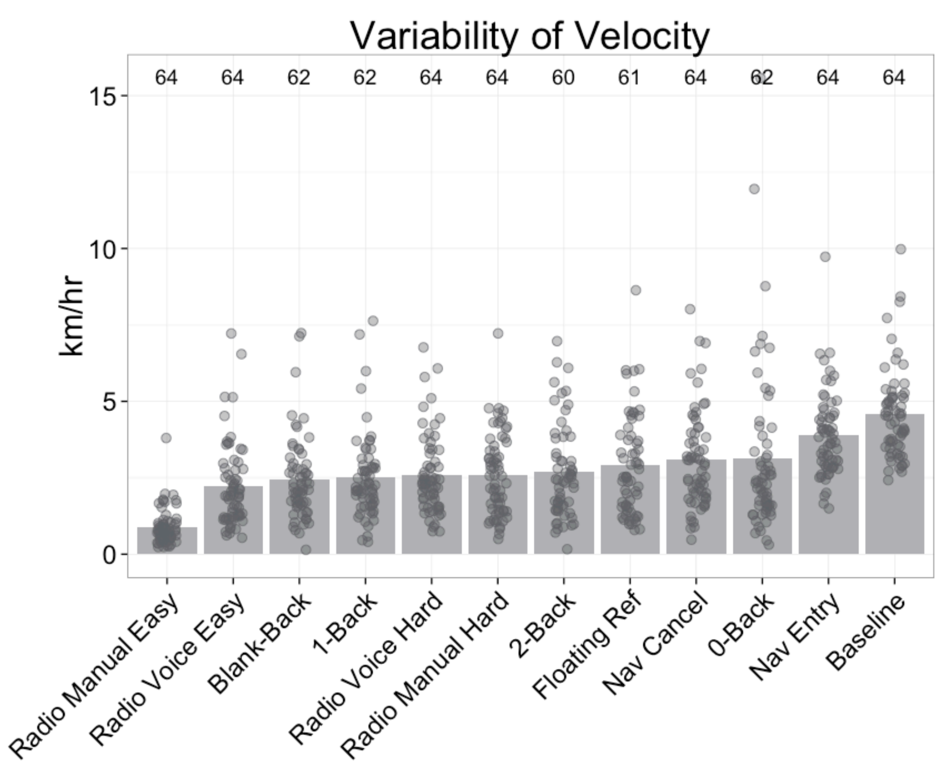


Figure 47: Variability (SD) of velocity with bars representing means and solid circles individual participants.

Consistent with the results of Study 1, velocity was least variable during the Radio Manual Easy task, and most variable during the Navigation Entry and Baseline periods. While the variability of vehicle speed during these tasks differed by as much as a factor of 4, the variability of speed control during the other tasks periods fell within a relatively narrower range. Again, it should be noted that the Radio Manual Easy task also had the shortest average task completion time, while Navigation Entry had the longest, and that task length can partially confound measures of standard deviation.

Statistical Comparison of Selected Tasks

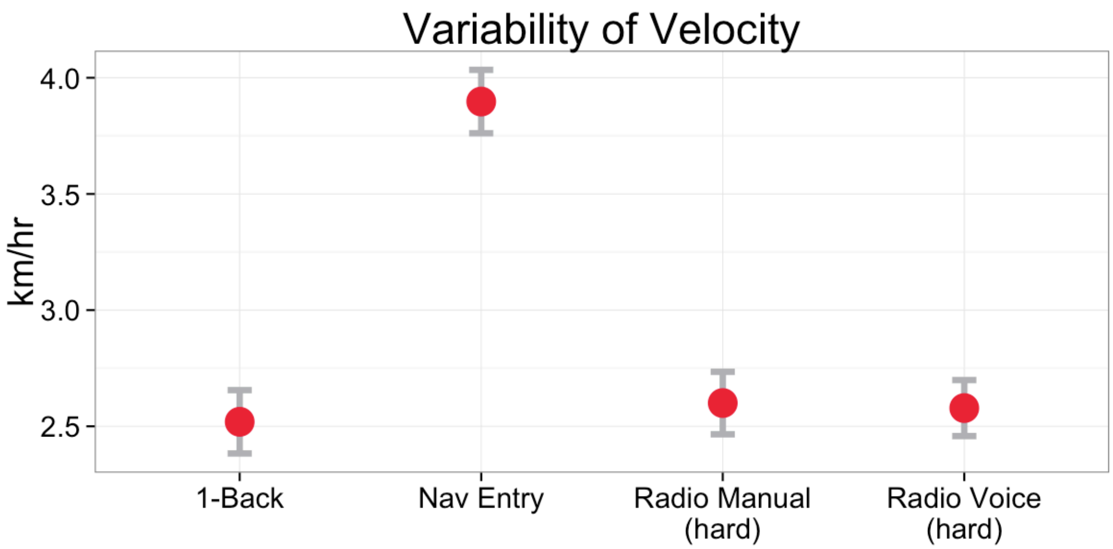


Figure 48: Variability (SD) of velocity across selected reference tasks.

Variability of velocity was significantly different across the four reference task periods ($X^2 = 43.6$, $p < .001$, Friedman test), and this effect was driven entirely by the Navigation Entry task, which had substantially greater speed variability than the other three tasks ($p < .001$ in comparison to each).

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 24: Mean and (SD) of variability of velocity by task, broken down by age group and gender.

Combined	20-24	25-39	40-54	55+	(all)
Baseline	4.70 (1.6)	4.77 (1.7)	4.55 (1.7)	4.30 (1.0)	4.58 (1.5)
Nav Entry	3.66 (1.5)	4.18 (2.0)	3.77 (0.9)	3.98 (1.0)	3.90 (1.4)
Nav Cancel	3.21 (1.2)	3.16 (1.9)	2.74 (1.1)	3.24 (2.0)	3.09 (1.6)
Radio Manual Easy	1.00 (0.8)	0.75 (0.3)	0.73 (0.4)	1.06 (0.6)	0.89 (0.6)
Radio Manual Hard	2.94 (1.6)	2.71 (1.2)	2.30 (1.4)	2.45 (1.0)	2.60 (1.3)
Radio Voice Easy	1.72 (1.1)	2.21 (1.1)	2.11 (1.1)	2.92 (1.9)	2.24 (1.4)
Radio Voice Hard	2.36 (1.2)	3.12 (1.5)	2.73 (1.3)	2.11 (1.0)	2.58 (1.3)
N-Back Reference	2.65 (1.1)	2.92 (1.4)	2.79 (1.6)	3.22 (2.3)	2.91 (1.6)
Blank-Back	2.21 (1.5)	2.81 (1.4)	2.04 (0.8)	2.75 (1.7)	2.46 (1.4)
0-Back	2.18 (1.6)	4.52 (3.5)	2.62 (2.1)	3.04 (2.7)	3.12 (2.7)
1-Back	1.80 (0.8)	3.18 (1.6)	3.01 (1.7)	2.01 (0.8)	2.52 (1.4)
2-Back	2.68 (1.8)	2.66 (1.3)	2.47 (1.4)	3.11 (1.8)	2.73 (1.6)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	4.78 (1.8)	3.54 (0.8)	4.63 (0.6)	4.26 (0.8)	4.30 (1.2)
Nav Entry	3.37 (1.5)	3.76 (1.5)	4.03 (1.0)	3.84 (0.9)	3.75 (1.2)
Nav Cancel	3.22 (1.2)	2.91 (2.1)	3.14 (0.8)	2.86 (2.1)	3.03 (1.6)
Radio Manual Easy	1.17 (1.1)	0.61 (0.2)	0.78 (0.5)	1.24 (0.7)	0.95 (0.7)
Radio Manual Hard	2.74 (1.2)	2.35 (1.1)	2.36 (1.6)	2.73 (1.1)	2.54 (1.2)
Radio Voice Easy	1.58 (1.0)	2.04 (0.6)	2.05 (1.0)	2.85 (2.5)	2.13 (1.5)
Radio Voice Hard	2.32 (1.6)	2.44 (1.1)	2.32 (0.9)	2.00 (1.1)	2.27 (1.1)
N-Back Reference	2.43 (1.1)	2.53 (1.6)	3.01 (1.5)	2.50 (1.3)	2.63 (1.4)
Blank-Back	1.60 (1.0)	3.16 (1.9)	2.36 (0.8)	2.66 (2.1)	2.50 (1.6)
0-Back	2.26 (1.2)	4.72 (4.8)	2.64 (1.8)	2.31 (1.5)	3.03 (2.9)
1-Back	1.55 (0.9)	3.00 (1.3)	3.17 (1.8)	1.80 (0.8)	2.44 (1.4)
2-Back	2.74 (1.8)	2.52 (0.6)	2.75 (1.6)	2.87 (1.8)	2.72 (1.4)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	4.63 (1.4)	6.01 (1.4)	4.46 (2.3)	4.34 (1.1)	4.86 (1.7)
Nav Entry	3.94 (1.5)	4.61 (2.4)	3.51 (0.8)	4.11 (1.0)	4.04 (1.5)
Nav Cancel	3.21 (1.3)	3.41 (1.8)	2.34 (1.3)	3.62 (2.0)	3.15 (1.6)
Radio Manual Easy	0.84 (0.3)	0.90 (0.4)	0.69 (0.3)	0.89 (0.5)	0.83 (0.4)
Radio Manual Hard	3.14 (2.1)	3.08 (1.3)	2.24 (1.2)	2.17 (0.9)	2.66 (1.5)
Radio Voice Easy	1.87 (1.1)	2.37 (1.4)	2.18 (1.4)	3.00 (1.0)	2.35 (1.2)
Radio Voice Hard	2.40 (0.7)	3.79 (1.6)	3.13 (1.6)	2.22 (0.9)	2.89 (1.4)
N-Back Reference	2.82 (1.1)	3.27 (1.4)	2.58 (1.7)	3.94 (2.8)	3.15 (1.9)
Blank-Back	2.66 (1.6)	2.47 (0.5)	1.72 (0.8)	2.84 (1.2)	2.42 (1.1)
0-Back	2.12 (1.9)	4.32 (1.8)	2.60 (2.5)	3.77 (3.5)	3.20 (2.5)
1-Back	1.98 (0.6)	3.35 (1.9)	2.85 (1.6)	2.21 (0.8)	2.60 (1.4)
2-Back	2.64 (1.9)	2.79 (1.8)	2.18 (1.2)	3.36 (1.8)	2.74 (1.7)

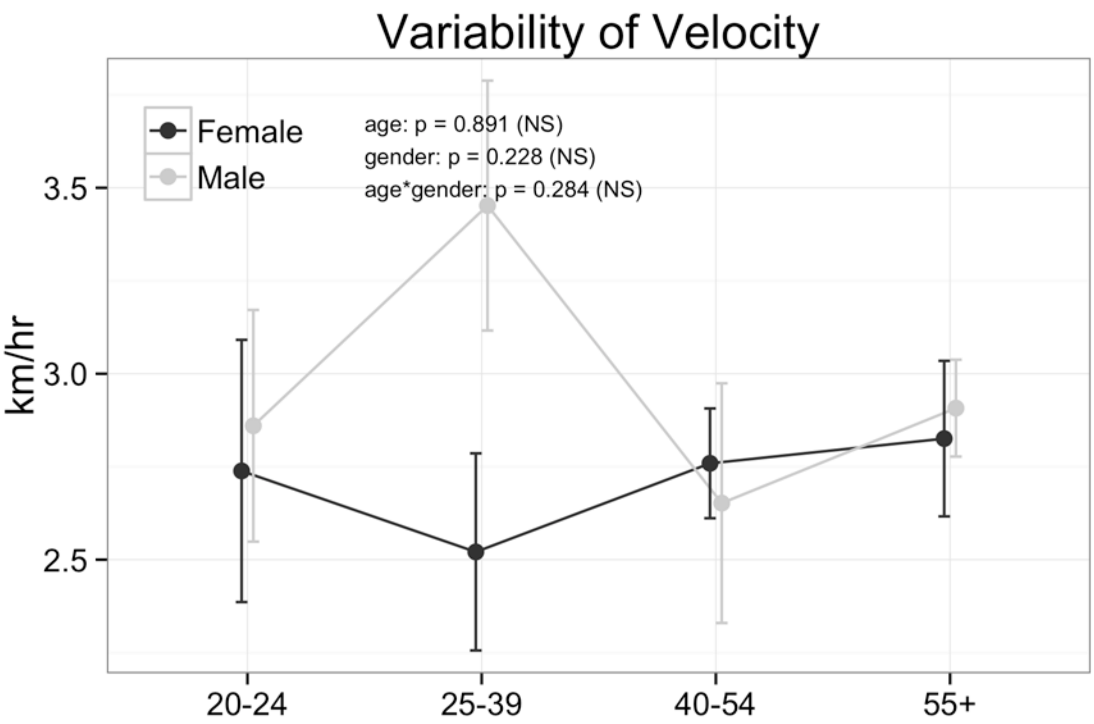


Figure 49: Variability (SD) of velocity across age groups and genders.

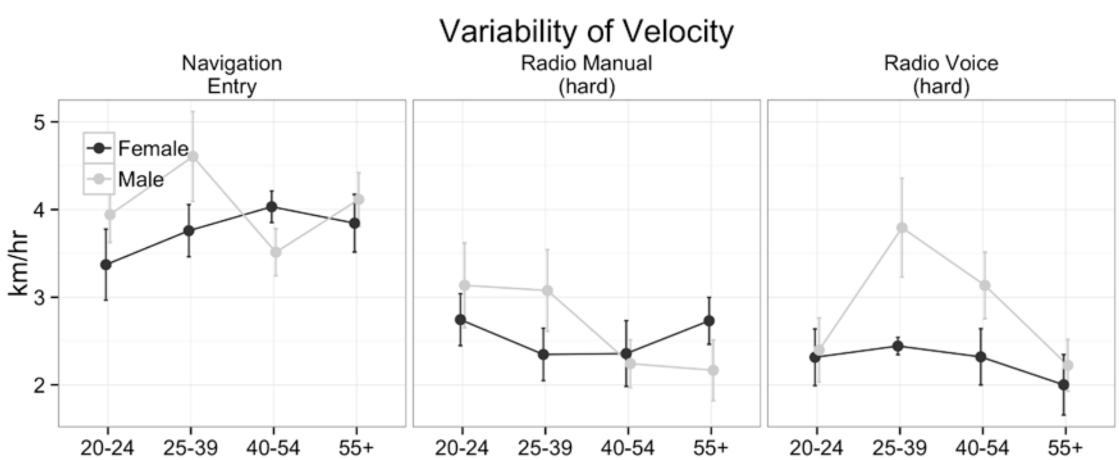


Figure 50: Variability (SD) of velocity by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Variability of velocity was not affected by age group or gender. No consistent patterns are apparent, aside from the fact that women appear to be nominally more consistent on this measure compared to men across age groups.

Acceleration Events

A minimum threshold of 0.1g (0.98m/s^2) and a temporal separation of 2 seconds between independent events were applied in defining acceleration events for this report. Following this metric, Table 25 displays acceleration events per minute for all DVI task periods. Acceleration events can be interpreted as an indicator of vehicular control: the greater the number of acceleration events, the less smoothly the driver was able to control the vehicle over the given driving period.

Table 25: Mean (and standard deviation) of acceleration events by training type.

	Structured Training	Self-Guided Training	(combined)
Baseline	0.93 (0.9)	1.00 (1.1)	0.96 (1.0)
Nav Entry	0.92 (1.2)	0.93 (1.1)	0.93 (1.1)
Nav Cancel	0.98 (1.6)	1.12 (1.6)	1.05 (1.6)
Radio Manual Easy	1.15 (3.0)	1.18 (2.9)	1.16 (2.9)
Radio Manual Hard	0.89 (1.1)	1.13 (1.8)	1.01 (1.5)
Radio Voice Easy	1.22 (2.5)	1.12 (1.9)	1.17 (2.2)
Radio Voice Hard	0.73 (1.8)	0.62 (1.2)	0.67 (1.5)

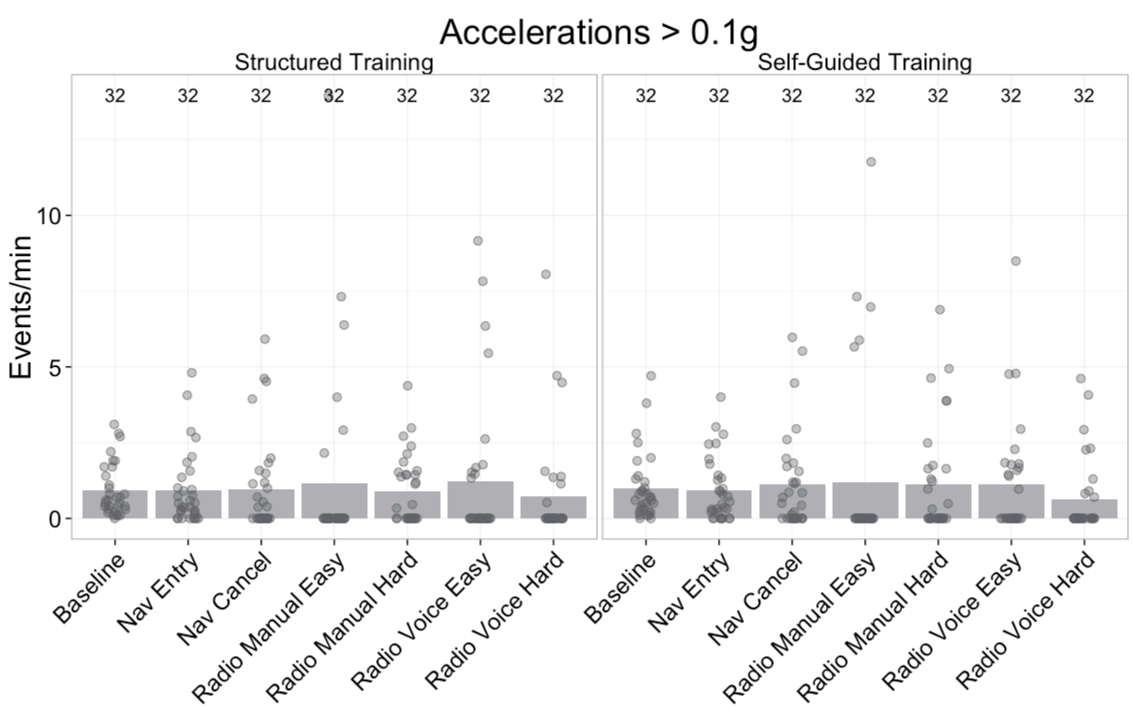


Figure 51: Acceleration events by task and training condition.

Acceleration events were not affected by training condition (p = .749).

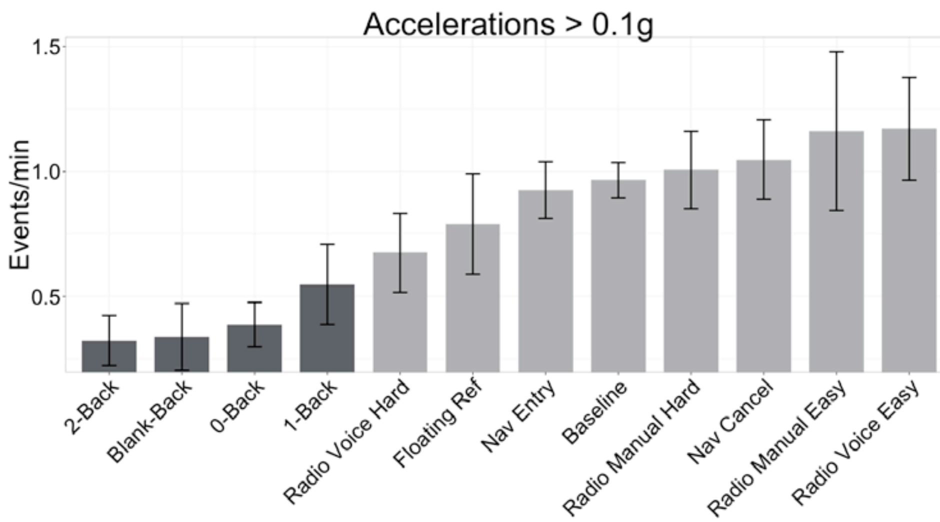


Figure 52: Acceleration events collapsed across training conditions showing means and adjusted standard errors. Levels of the n-back task are shaded in dark gray

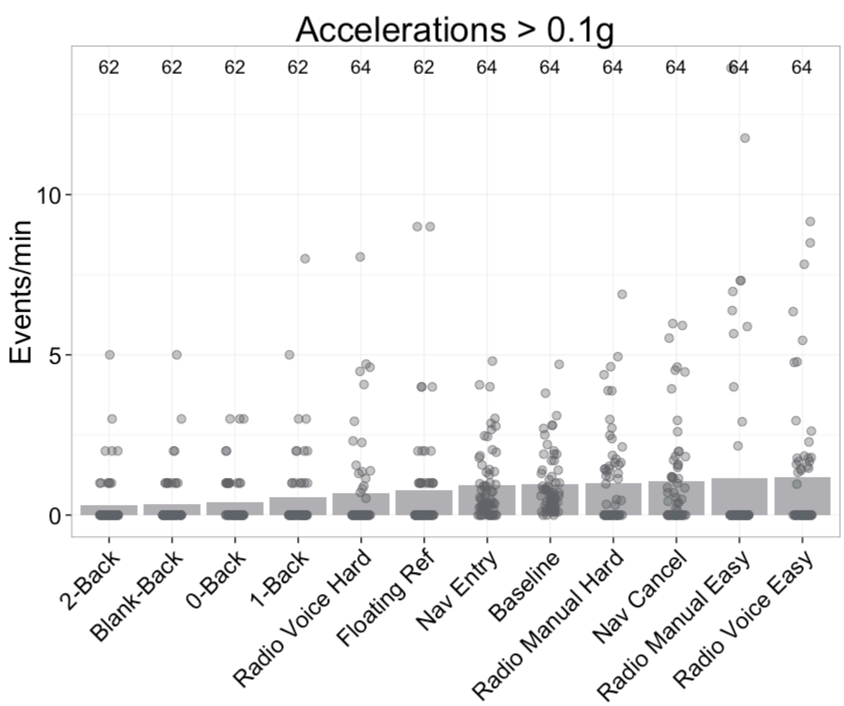


Figure 53: Acceleration events with bars representing means and solid circles individual participants.

Note that all of the n-back tasks cluster at the low end of the scale, exhibiting fewer acceleration events than the DVI tasks, or even the baseline period. This may indicate a compensatory effect on the part of the driver: as cognitive demand is increased (via the n-back task), the driver may compensate for his divided attention by focusing on the forward roadway, driving more conservatively, thus reducing acceleration events. The between-task patterning of acceleration events is broadly consistent with that observed in Study 1, though there was a nominally higher overall frequency of acceleration events in this sample.

Statistical Comparison of Selected Tasks

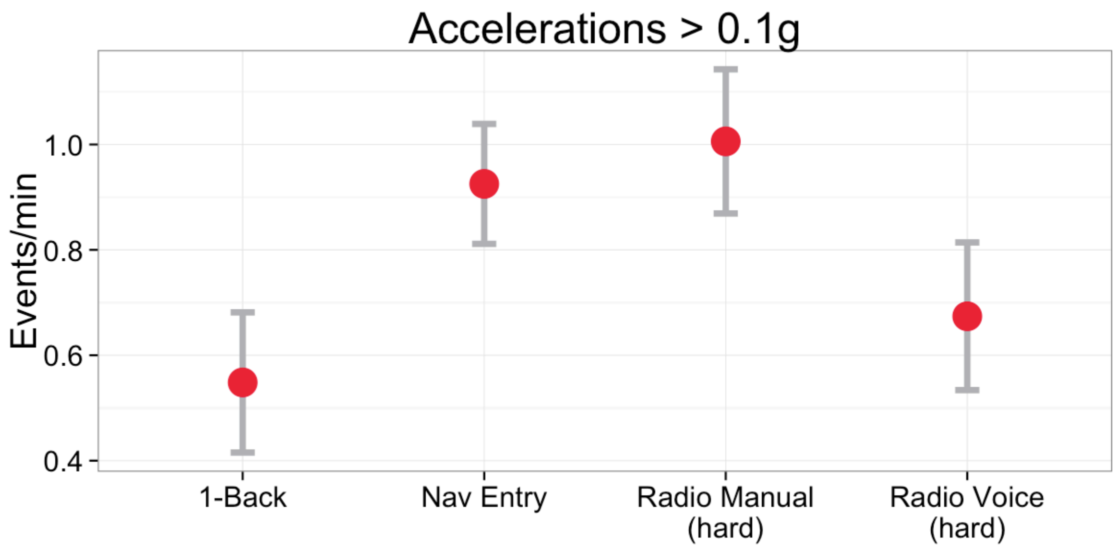


Figure 54: Acceleration events across selected reference tasks.

Acceleration event frequency varied significantly between the four reference tasks ($X^2 = 20.5$, $p < .001$, Friedman test). As reported above, the 1-back task period had the lowest acceleration event rate, which was significantly lower than the Navigation Entry and Radio Manual Hard rates ($p = .006$ and $p = .045$, respectively). Additionally, the Radio Voice Hard task had significantly fewer acceleration events than the equivalent Radio Manual Hard task ($p = .035$).

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 26: Mean and (SD) of acceleration events by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	0.99 (0.8)	1.37 (1.2)	0.79 (1.1)	0.71 (0.7)	0.96 (1.0)
Nav Entry	1.08 (1.3)	1.16 (1.2)	0.63 (0.9)	0.82 (1.1)	0.93 (1.1)
Nav Cancel	1.88 (2.1)	1.19 (1.7)	0.42 (0.7)	0.70 (1.2)	1.05 (1.6)
Radio Manual Easy	1.18 (2.6)	2.68 (4.7)	0.00 (0.0)	0.78 (1.9)	1.16 (2.9)
Radio Manual Hard	1.07 (1.8)	1.28 (1.7)	1.29 (1.5)	0.39 (0.6)	1.01 (1.5)
Radio Voice Easy	1.28 (2.0)	2.58 (3.4)	0.40 (0.7)	0.42 (0.8)	1.17 (2.2)
Radio Voice Hard	1.12 (2.3)	1.04 (1.6)	0.48 (1.2)	0.05 (0.2)	0.67 (1.5)
N-Back Reference	1.21 (2.4)	0.88 (1.4)	1.06 (2.4)	0.06 (0.2)	0.79 (1.8)
Blank-Back	0.43 (0.9)	0.12 (0.3)	0.19 (0.5)	0.62 (1.3)	0.34 (0.8)
0-Back	0.29 (0.6)	0.75 (1.2)	0.06 (0.2)	0.44 (0.6)	0.39 (0.8)
1-Back	0.50 (0.9)	0.94 (2.3)	0.38 (0.9)	0.38 (0.7)	0.55 (1.4)
2-Back	0.43 (0.9)	0.56 (1.3)	0.19 (0.4)	0.12 (0.5)	0.32 (0.9)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	0.70 (0.7)	1.00 (1.2)	1.01 (1.5)	0.58 (0.7)	0.82 (1.0)
Nav Entry	1.35 (1.7)	0.73 (0.7)	0.70 (0.9)	0.57 (1.0)	0.84 (1.1)
Nav Cancel	1.36 (1.8)	1.03 (1.9)	0.84 (0.9)	0.98 (1.6)	1.05 (1.5)
Radio Manual Easy	0.91 (2.6)	1.47 (4.2)	0.00 (0.0)	1.07 (2.3)	0.86 (2.6)
Radio Manual Hard	1.06 (1.0)	0.47 (0.8)	1.84 (1.8)	0.37 (0.5)	0.94 (1.2)
Radio Voice Easy	0.49 (1.0)	1.19 (2.2)	0.39 (0.7)	0.72 (1.1)	0.70 (1.3)
Radio Voice Hard	0.17 (0.5)	0.29 (0.5)	0.58 (1.6)	0.00 (0.0)	0.26 (0.9)
N-Back Reference	2.00 (3.5)	0.38 (0.7)	1.50 (3.1)	0.00 (0.0)	0.90 (2.3)
Blank-Back	0.50 (1.2)	0.00 (0.0)	0.38 (0.7)	0.25 (0.5)	0.27 (0.7)
0-Back	0.00 (0.0)	0.62 (1.1)	0.12 (0.4)	0.25 (0.5)	0.27 (0.6)
1-Back	0.33 (0.5)	0.62 (1.8)	0.38 (1.1)	0.62 (0.9)	0.50 (1.1)
2-Back	0.17 (0.4)	0.38 (0.7)	0.12 (0.4)	0.25 (0.7)	0.23 (0.6)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	1.29 (0.9)	1.74 (1.2)	0.56 (0.3)	0.84 (0.8)	1.11 (0.9)
Nav Entry	0.82 (0.9)	1.59 (1.4)	0.56 (0.9)	1.08 (1.2)	1.01 (1.2)
Nav Cancel	2.40 (2.4)	1.36 (1.5)	0.00 (0.0)	0.41 (0.8)	1.04 (1.7)
Radio Manual Easy	1.44 (2.7)	3.89 (5.1)	0.00 (0.0)	0.50 (1.4)	1.46 (3.2)
Radio Manual Hard	1.07 (2.4)	2.10 (2.0)	0.74 (1.0)	0.40 (0.7)	1.07 (1.7)
Radio Voice Easy	2.06 (2.5)	3.97 (3.9)	0.41 (0.8)	0.12 (0.3)	1.64 (2.7)
Radio Voice Hard	2.08 (2.9)	1.78 (1.9)	0.39 (0.6)	0.10 (0.3)	1.09 (1.9)
N-Back Reference	0.62 (0.9)	1.38 (1.7)	0.62 (1.4)	0.12 (0.4)	0.69 (1.2)
Blank-Back	0.38 (0.7)	0.25 (0.5)	0.00 (0.0)	1.00 (1.7)	0.41 (1.0)
0-Back	0.50 (0.8)	0.88 (1.4)	0.00 (0.0)	0.62 (0.7)	0.50 (0.9)
1-Back	0.62 (1.1)	1.25 (2.8)	0.38 (0.7)	0.12 (0.4)	0.59 (1.5)
2-Back	0.62 (1.2)	0.75 (1.8)	0.25 (0.5)	0.00 (0.0)	0.41 (1.1)

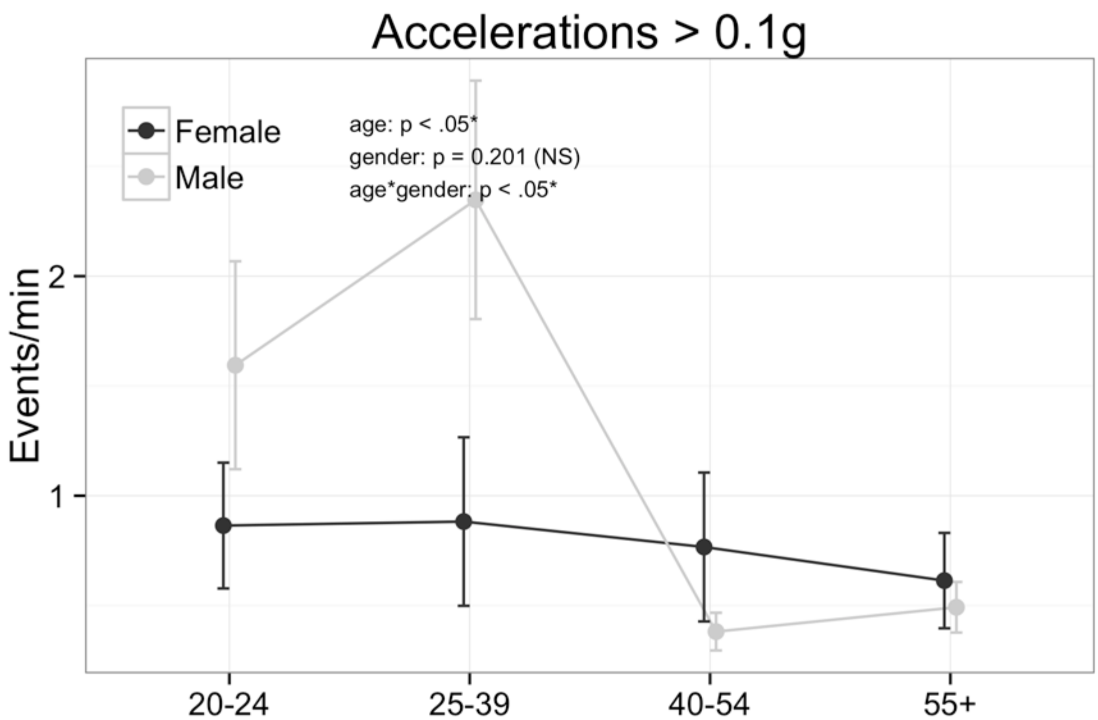


Figure 55: Acceleration events across age groups and genders.

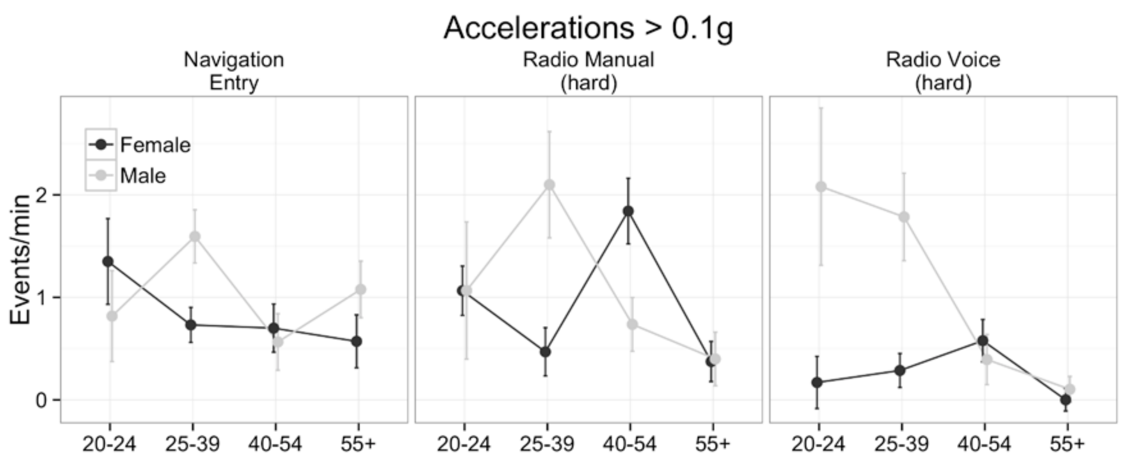


Figure 56: Acceleration events by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing by age and gender, a significant effect of age is evident (p = .029), as well as an age * gender interaction (p < .05). This effect is largely driven by older men, who have fewer acceleration events after age 40. In contrast, the driving characteristics of women in regards to acceleration events is generally consistent across age groups.

Acceleration Events — Extended Detail

Table 27: Mean count (& standard deviation) of acceleration events ≥ 0.10 g by age group (N=64)

	20-24	25-39	40-54	55+	Event Count
≥ 0.10 g	1.10±2.28	1.42±2.93	0.76±1.95	0.79±2.12	1107
≥ 0.15 g	0.11±0.49	0.14±0.50	0.06±0.24	0.10±0.69	112
≥ 0.20 g	0.01±0.10	0.02±0.16	0.004±0.06	0.03±0.43	18
≥ 0.30 g	N/A	0.004±0.06		0.007±0.12	3

Table 27 summarizes unidirectional acceleration events across all subjects by age group. No events exceeding a threshold of 0.35 g were encountered in the study data set. Only three independent acceleration events in excess of 0.3 g were found. Upon manual review of these events, two can be attributed to a driver who briefly exited the highway onto a local roadway after being instructed to perform the navigation cancel task. Upon exiting the roadway, the subject executed a series of low speed maneuvers to return to the onramp of the highway. Therefore, only one acceleration event in excess of 0.3 g can be considered a genuine independent event in the context of this assessment. This event was caused by a subject's steering inputs while performing a DVI task (the manual radio tuning reference task – manual radio hard). A total of 18 acceleration events in excess of 0.2 g were recorded. Seven of these can be attributed to the subject who exited the highway. Additionally, since accelerations are counted cumulatively, the lone 0.3 g acceleration from the Radio Manual Hard task is also counted as a 0.2 g acceleration.

Revised acceleration event totals that account for these issues are presented in Table 28 (raw totals that do not remove accelerations due to erroneous task performance are shown in parentheses). With these events removed, there was a total of 1 event greater than 0.3 g, and 10 events greater than 0.2 g. Four of the 0.2 g events occurred during baseline driving. On average, subjects experienced 9.1 minutes of baseline driving and 8.4 minutes of task period driving. In other words, driving time was split almost equally between task and baseline periods, as were acceleration events greater than 0.2 g. The distribution of acceleration events between task and baseline periods is consistent with the results of Study 1, once again suggesting that the occurrence of acceleration events is largely dependent on time spent driving, and not task demands, specifically.

Table 28: Cumulative count of acceleration events > 2.0 g by task type (uncorrected totals in parentheses, see discussion above).

	≥ 0.20 g	≥ 0.30 g
Baseline	4	
Radio voice activation (easy)	2	
Radio voice activation (hard)		
Radio manual input (easy)		
Radio manual input (hard)	2 (3)	1
Navigation entry	2	
Navigation cancel	0 (7)	0 (2)

Table 29: Details of major acceleration events. Events resulting from one subject's departure from the highway are highlighted in italics.

Acceleration	Direction	Magnitude	Subject	Task	Reason
≥ 0.30g	Lateral	0.31 g	Subject 30 Female, 55+ Untrained	Navigation cancel	Subject pulled into driveway for turning maneuver
	Lateral	0.33 g	Subject 30 Female, 55+ Untrained	Navigation cancel	Abrupt merge onto highway ramp

	Lateral	0.35 g	Subject 71 Male, 25-39 Untrained	Radio manual input (hard)	Loss of lateral control while looking at HMI during task completion
≥ 0.20g	Lateral	0.21 g	Subject 13 Male, 20-24 Untrained	Radio voice input (easy)	Damaged road surface
	Lateral	0.21 g	Subject 15 Male, 25-93 Trained	Baseline	Passing maneuver
	Lateral	0.22 g	Subject 21 Male, 20-24 Untrained	0-Back verbal	Passing maneuver
	Lateral	0.21 g	Subject 21 Male, 20-24 Untrained	1-Back self-paced	Lane change
	Lateral	0.20 g	Subject 23 Male, 20-24 Untrained	Radio voice input (easy)	Damaged road surface
	Lateral	0.22 g	Subject 27 Female, 25- Subject 39 Untrained	Navigation entry	Loss of lateral control while looking at HMI during task completion
	Lateral	0.21 g	Subject 27 Female, 25- Subject 39 Untrained	Radio manual input (hard)	Loss of lateral control while looking at HMI during task completion
	Both	*~ 0.2 g*	*Subject 30* *Female, 55+* *Untrained*	*Navigation cancel*	*Several accel. events during turning maneuver*

Lateral	0.21 g	Subject 44 Male, 40-54 Trained	1-Back fixed-paced	Damaged road surface
Longitudinal	-0.21 g	Subject 48 Female, 55+ Untrained	Baseline	Abrupt braking due to merging traffic ahead
Longitudinal	-0.24 g	Subject 60 Female, 55+ Trained	Baseline	Braking during passing maneuver
Longitudinal	-0.2 g	Subject 61 Female, 20-Subject 24 Trained	Baseline	Abrupt braking due to slow traffic ahead
Both	0.23 g	Subject 65 Female, 40-Subject 54 Trained	Navigation entry	Damaged road surface
Lateral	0.21 g	Subject 69 Male, 25-39 Untrained	1-Back fixed-paced	Damaged road surface
Lateral	0.20 g	Subject 71 Male, 25-39 Untrained	Radio manual input (hard)	Loss of lateral control while looking at HMI during task completion

Variability of Steering Wheel Angle

Increases in variability (standard deviation) in steering wheel angle is commonly related to reduced lateral control and associated with a driver's need to concurrently manage the added workload of secondary activates (Östlund et al., 2004). Under normal driving conditions, minor steering wheel corrections are made to adjust vehicle heading for variations in roadway conditions (Liu, Schreiner, & Dinges, 1999). These variations can be looked at using a number of different methods including the standard deviation of wheel angle, and counts of minor (small) wheel reversals and major (large) wheel reversals. In situations of increased cognitive workload, the number of minor steering wheel adjustments tend to increase, while secondary activates that involve visual attention demands often impact large reversals (Östlund, et al., 2005).

Table 30: Mean (and standard deviation) of variability of wheel angle by training type.

	Structured Training	Self-Guided Training	(combined)
Baseline	1.91 (0.2)	1.90 (0.3)	1.90 (0.2)
Nav Entry	1.86 (0.4)	1.86 (0.4)	1.86 (0.4)
Nav Cancel	1.63 (0.5)	1.58 (0.4)	1.60 (0.4)
Radio Manual Easy	1.20 (0.5)	1.15 (0.3)	1.17 (0.4)
Radio Manual Hard	1.58 (0.3)	1.47 (0.5)	1.53 (0.4)
Radio Voice Easy	1.32 (0.4)	1.36 (0.4)	1.34 (0.4)
Radio Voice Hard	1.31 (0.5)	1.45 (0.4)	1.38 (0.5)

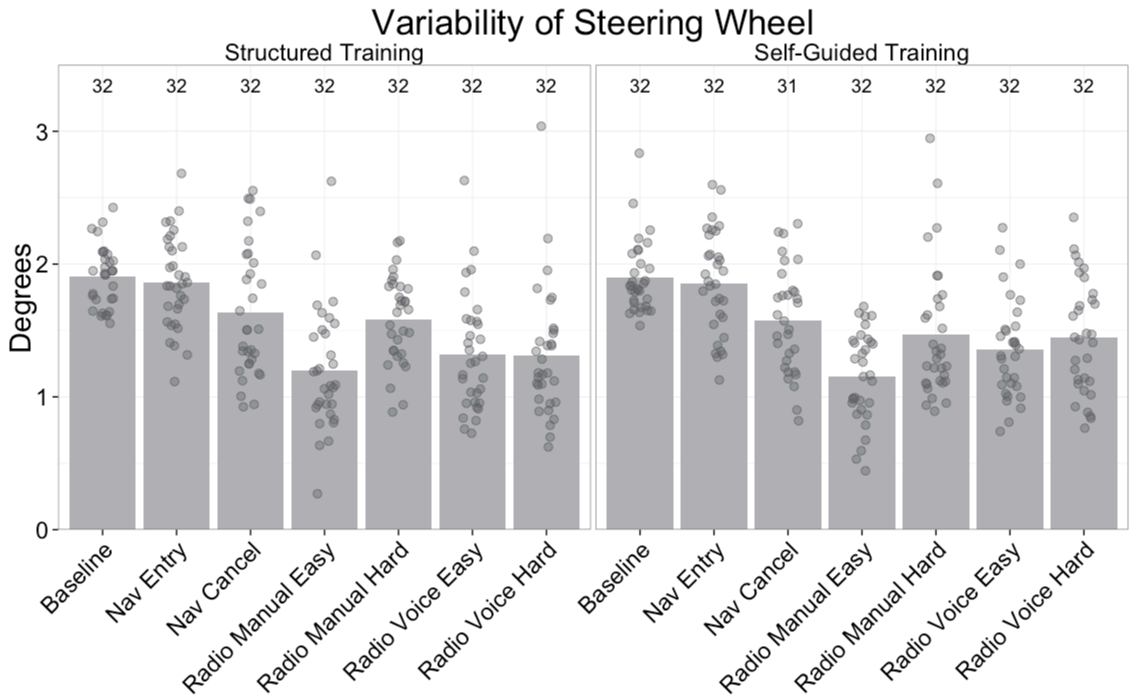


Figure 57: Variability (SD) of wheel angle by task and training condition.

As with other primary measures considered, variability of wheel angle was not affected by training condition (p = .902).

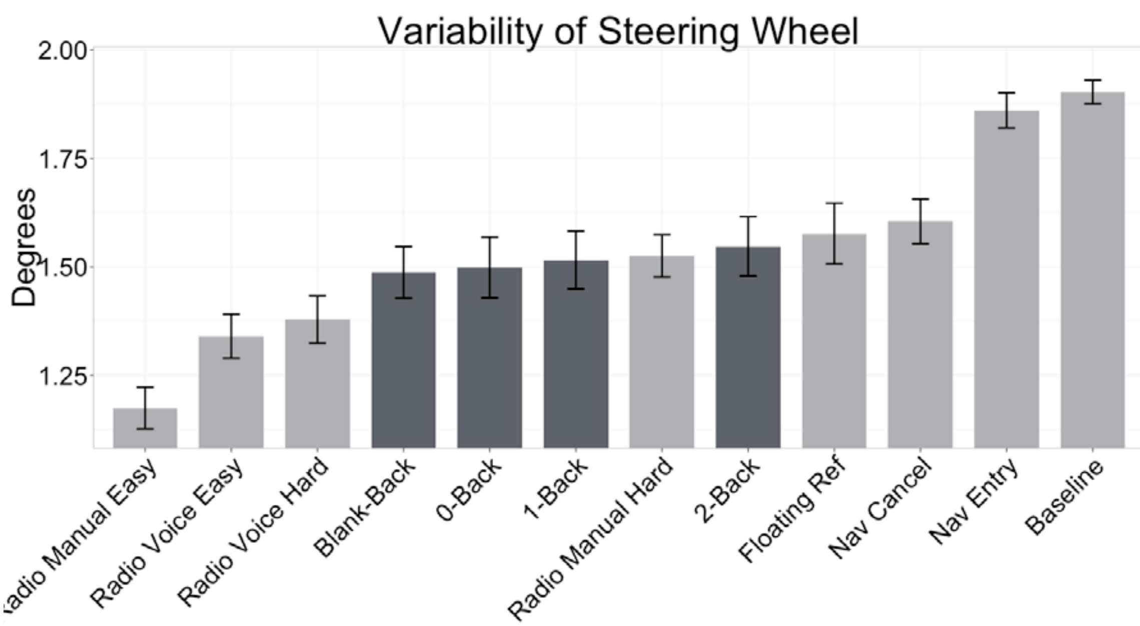


Figure 58: Variability (SD) of steering wheel angle showing means and adjusted standard errors. Levels of the n-back task are shaded in dark gray.

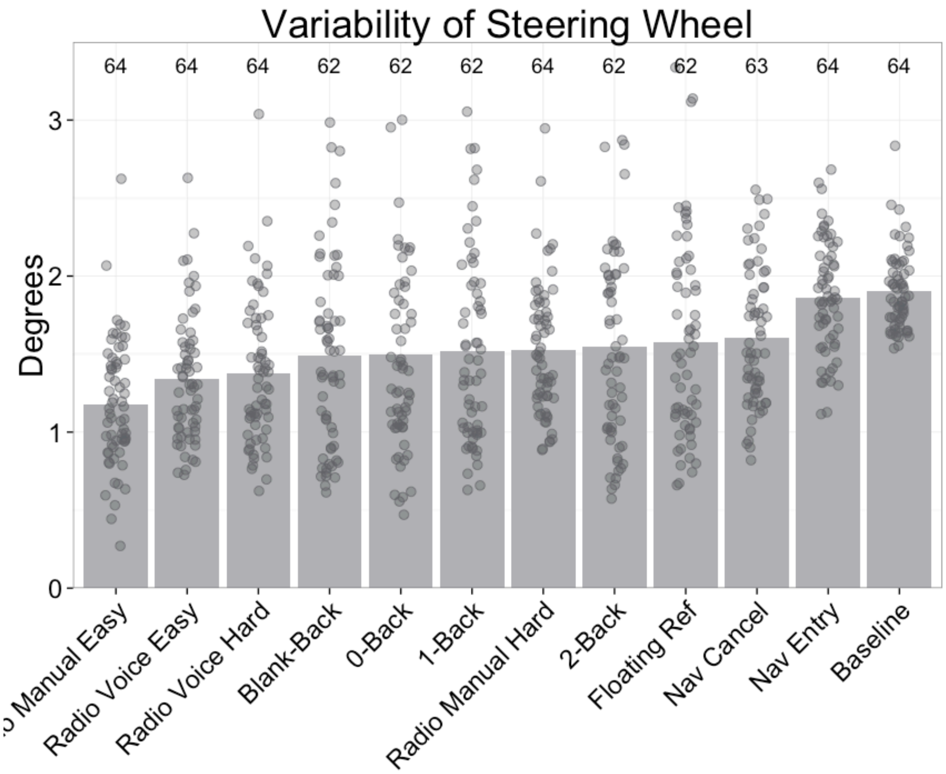


Figure 59: Variability (SD) of wheel angle with bars representing mean SD values and solid circles individual participants.

Steering wheel angle ranged from a minimum deviation of 1.17º during the Radio Manual Easy task and a maximum deviation of 1.90º during the Baseline period (closely followed by 1.86º in the Navigation Entry task). This pattern is broadly consistent with the results of Study 1, in which the Baseline and Navigation entry had the largest wheel position variability (2.23º and 2.25º, respectively), while the Radio Manual Easy task had the smallest variability of any DVI task (1.23º). The present data differ from our previous results in that wheel variability is slightly lower here, and the levels of the n-back task, which clustered at the low end of variability in the previous report, are somewhat more broadly distributed here.

Statistical Comparison of Selected Tasks

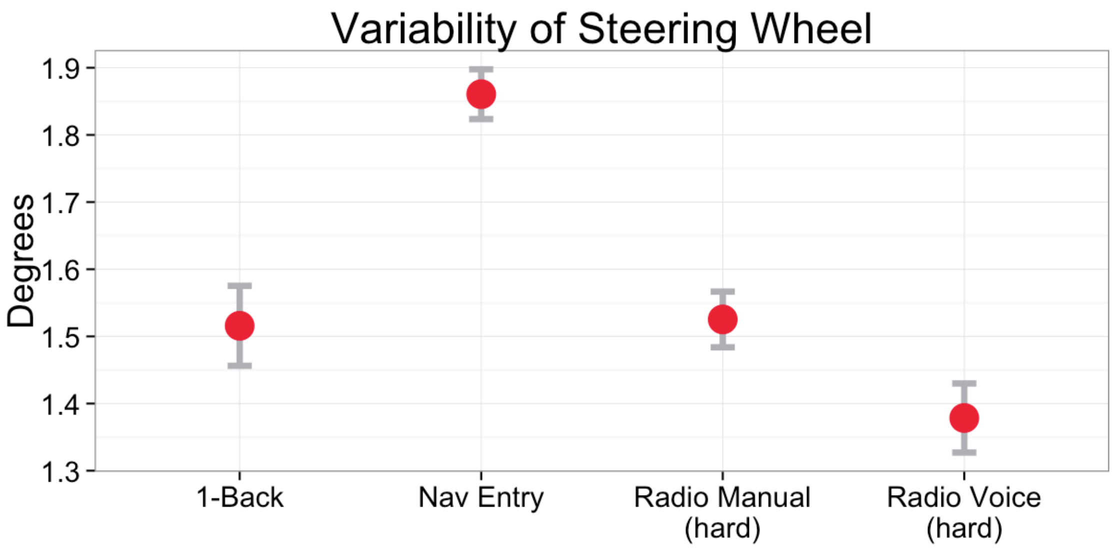


Figure 60: Variability (SD) of wheel angle across selected reference tasks.

Variability of wheel position varied significantly across the four reference tasks (X^2 = 32.4, p < .001, Friedman test). The difference was mainly driven by the Navigation Entry task, which had significantly higher wheel position variability compared to each of the other three tasks (p < .01 for all three comparisons, Wilcoxon signed rank tests). The Radio Manual Hard task had significantly higher variability compared to the equivalent Radio Voice Hard task (p = .040).

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 31: Mean and (SD) of variability of wheel angle by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	1.96 (0.2)	1.90 (0.3)	1.90 (0.3)	1.85 (0.2)	1.90 (0.2)
Nav Entry	1.83 (0.3)	1.86 (0.4)	1.91 (0.4)	1.85 (0.4)	1.86 (0.4)
Nav Cancel	1.77 (0.5)	1.62 (0.3)	1.52 (0.5)	1.50 (0.4)	1.60 (0.4)
Radio Manual Easy	1.09 (0.4)	1.19 (0.4)	1.14 (0.4)	1.28 (0.5)	1.17 (0.4)
Radio Manual Hard	1.45 (0.4)	1.59 (0.5)	1.61 (0.4)	1.45 (0.4)	1.53 (0.4)
Radio Voice Easy	1.34 (0.3)	1.34 (0.5)	1.38 (0.4)	1.30 (0.3)	1.34 (0.4)
Radio Voice Hard	1.38 (0.6)	1.36 (0.4)	1.37 (0.4)	1.40 (0.4)	1.38 (0.5)
N-Back Reference	1.67 (0.7)	1.63 (0.7)	1.63 (0.7)	1.39 (0.5)	1.58 (0.6)
Blank-Back	1.38 (0.7)	1.61 (0.7)	1.36 (0.5)	1.59 (0.6)	1.49 (0.6)
0-Back	1.48 (0.7)	1.72 (0.9)	1.25 (0.4)	1.54 (0.6)	1.50 (0.7)
1-Back	1.41 (0.5)	1.66 (0.7)	1.45 (0.6)	1.54 (0.6)	1.52 (0.6)
2-Back	1.50 (0.8)	1.71 (0.6)	1.45 (0.6)	1.52 (0.6)	1.55 (0.7)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	1.88 (0.3)	1.88 (0.3)	1.97 (0.4)	1.83 (0.2)	1.89 (0.3)
Nav Entry	1.87 (0.4)	1.69 (0.4)	1.98 (0.5)	1.74 (0.3)	1.82 (0.4)
Nav Cancel	1.84 (0.4)	1.60 (0.2)	1.79 (0.4)	1.50 (0.5)	1.69 (0.4)
Radio Manual Easy	1.04 (0.3)	1.13 (0.3)	1.14 (0.4)	1.25 (0.3)	1.14 (0.3)
Radio Manual Hard	1.41 (0.3)	1.50 (0.3)	1.62 (0.4)	1.42 (0.3)	1.49 (0.3)
Radio Voice Easy	1.15 (0.2)	1.18 (0.4)	1.24 (0.3)	1.32 (0.4)	1.22 (0.3)
Radio Voice Hard	1.11 (0.3)	1.22 (0.3)	1.21 (0.4)	1.41 (0.5)	1.24 (0.4)
N-Back Reference	1.81 (0.9)	1.26 (0.5)	1.75 (0.8)	1.45 (0.6)	1.55 (0.7)
Blank-Back	1.25 (0.6)	1.52 (0.6)	1.48 (0.5)	1.36 (0.4)	1.41 (0.5)
0-Back	1.18 (0.3)	1.63 (0.9)	1.36 (0.5)	1.32 (0.7)	1.39 (0.7)
1-Back	1.26 (0.5)	1.70 (0.8)	1.28 (0.7)	1.44 (0.6)	1.43 (0.6)
2-Back	1.67 (0.6)	1.67 (0.7)	1.53 (0.7)	1.51 (0.7)	1.59 (0.6)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	2.04 (0.2)	1.93 (0.3)	1.82 (0.2)	1.88 (0.2)	1.92 (0.2)
Nav Entry	1.79 (0.3)	2.03 (0.4)	1.83 (0.3)	1.95 (0.4)	1.90 (0.4)
Nav Cancel	1.70 (0.5)	1.64 (0.4)	1.25 (0.5)	1.51 (0.5)	1.52 (0.5)
Radio Manual Easy	1.14 (0.4)	1.25 (0.5)	1.13 (0.3)	1.31 (0.6)	1.21 (0.5)
Radio Manual Hard	1.48 (0.5)	1.69 (0.7)	1.60 (0.4)	1.48 (0.4)	1.56 (0.5)
Radio Voice Easy	1.52 (0.4)	1.50 (0.6)	1.52 (0.5)	1.29 (0.3)	1.46 (0.5)
Radio Voice Hard	1.66 (0.7)	1.50 (0.5)	1.52 (0.4)	1.40 (0.3)	1.52 (0.5)
N-Back Reference	1.57 (0.5)	2.00 (0.7)	1.52 (0.5)	1.32 (0.5)	1.60 (0.6)
Blank-Back	1.48 (0.7)	1.71 (0.7)	1.24 (0.4)	1.81 (0.7)	1.56 (0.7)
0-Back	1.70 (0.9)	1.81 (0.9)	1.14 (0.2)	1.77 (0.5)	1.60 (0.7)
1-Back	1.52 (0.5)	1.61 (0.7)	1.61 (0.5)	1.63 (0.6)	1.60 (0.6)
2-Back	1.37 (1.0)	1.75 (0.6)	1.38 (0.4)	1.54 (0.6)	1.51 (0.7)

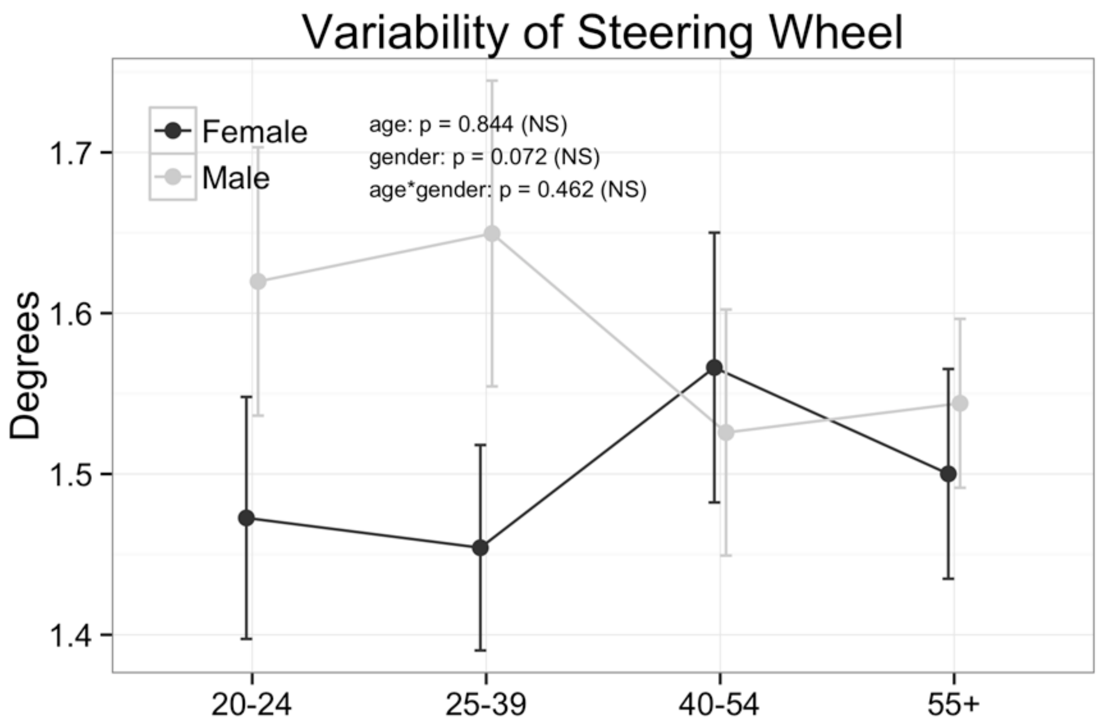


Figure 61: Variability (SD) of wheel angle across age groups and genders.

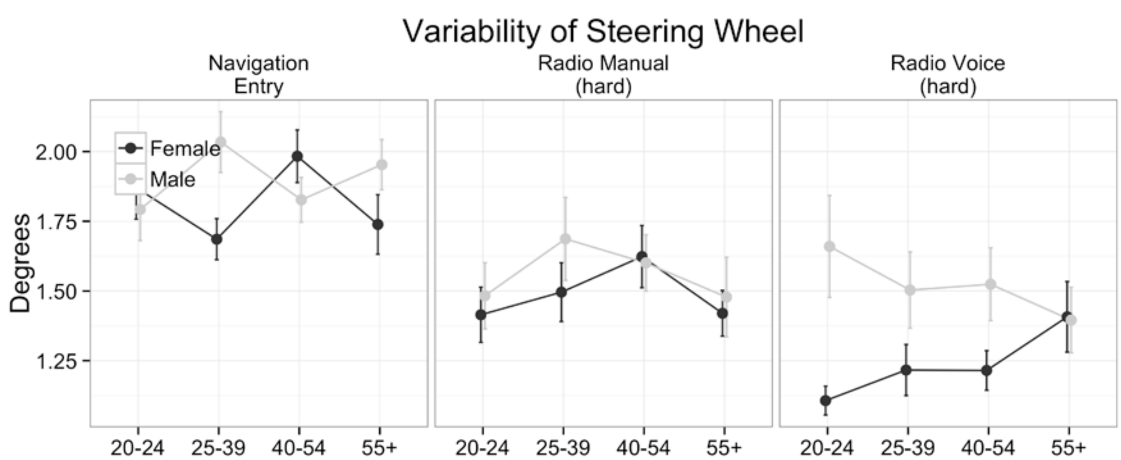


Figure 62: Variability (SD) of wheel angle by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing variability of wheel angle by age and gender, there were no statistically significant main effects. There was a borderline trend suggestive of a gender influence (p = .072), that may be driven by an apparent age * gender interaction in the Radio Voice Hard task.

Minor Steering Wheel Reversal Rate

Minor (small) steering wheel reversals (SWR) were counted and classified using a filter with a 0.1^{o} gap size (see Methods for details). Like the mean velocity and variability of velocity measures described above, an increase in this measure may reflect the impact of task demands on vehicle control.

Table 32: Mean (and standard deviation) of minor SWR by training type.

	Structured Training	Self-Guided Training	(combined)
Baseline	73.42 (11.9)	74.68 (9.4)	74.05 (10.6)
Nav Entry	82.00 (12.1)	83.27 (11.8)	82.64 (11.9)
Nav Cancel	81.17 (14.5)	84.45 (12.1)	82.81 (13.3)
Radio Manual Easy	76.12 (13.7)	77.74 (18.1)	76.93 (15.9)
Radio Manual Hard	81.15 (12.0)	82.78 (9.5)	81.96 (10.8)
Radio Voice Easy	81.29 (14.2)	82.18 (14.0)	81.73 (14.0)
Radio Voice Hard	82.65 (13.6)	81.76 (10.6)	82.20 (12.1)

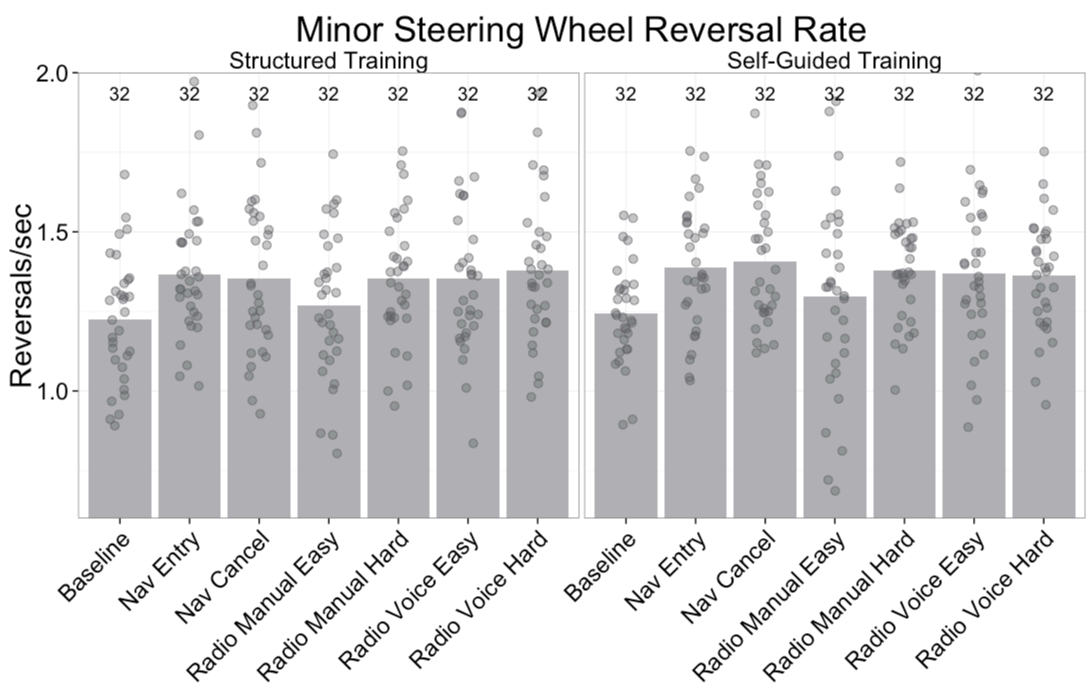


Figure 63: Minor SWR by task and training condition.

Minor SWR was not affected by training condition ($p = .510$). Once again, the pattern of mean results is fairly consistent between the two training conditions.

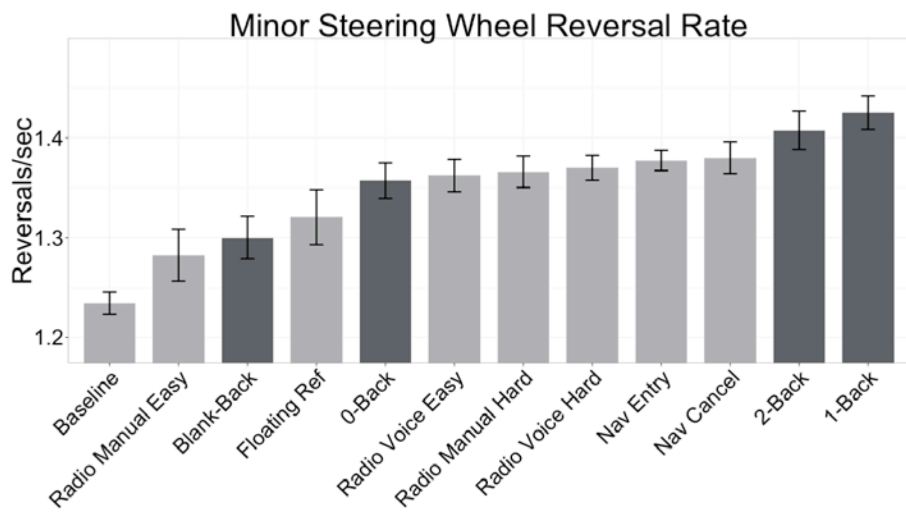


Figure 64: Minor SWR means and adjusted standard errors. Levels of the n-back task are shaded in dark gray.

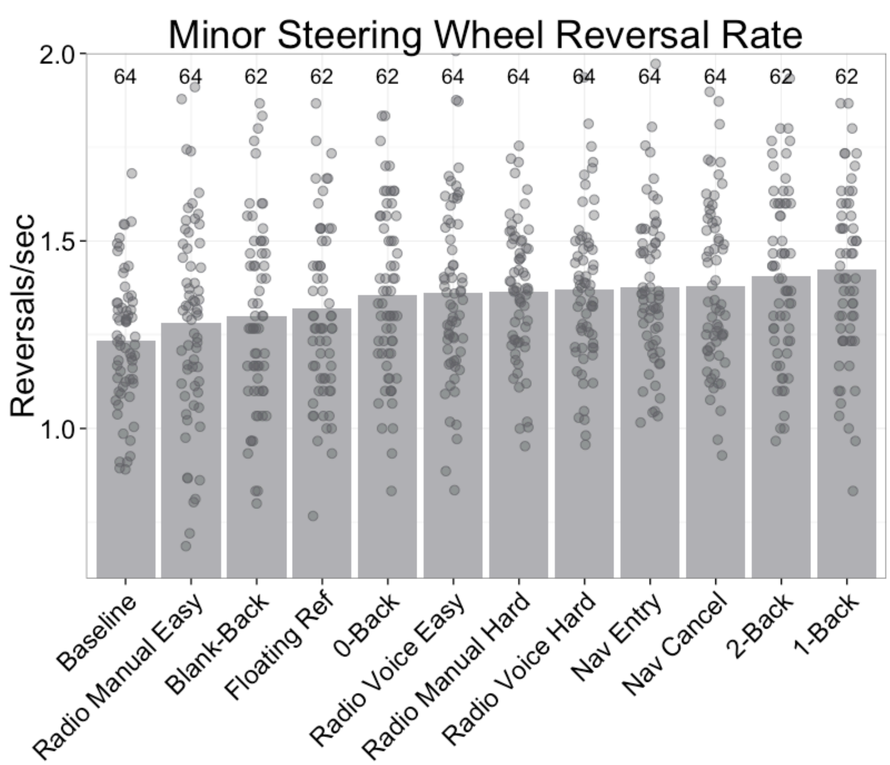


Figure 65: Minor SWR with bars representing means and solid circles individual participants.

Note that individual data points for the n-back periods in Figure 65 appear somewhat discontinuous (i.e. "stuttered" or separated) because these periods employed fixed task durations. Minor SWR increases from baseline driving during all of the task periods, with the no differentiation between DVI tasks on this measure, with the possible exception of the Radio Manual Easy tuning task which appears in this sample as intermediate between baseline and the other DVI tasks. Radio Manual Easy was not distinguishable from the other DVI tasks in Study 1, so this finding may not be particularly generalizable. The 1- and 2-back tasks again appear at the high end of the minor SWR distribution as they did in Study 1 and the 0-back and blank-back toward the lower end of the distribution. However, as indicated in the planned statistical comparisons below, minor SWR is not providing clear differentiation across task types, other than the observation that it does increase relative to baseline driving with drivers are engaged in a task.

Statistical Comparison of Selected Tasks

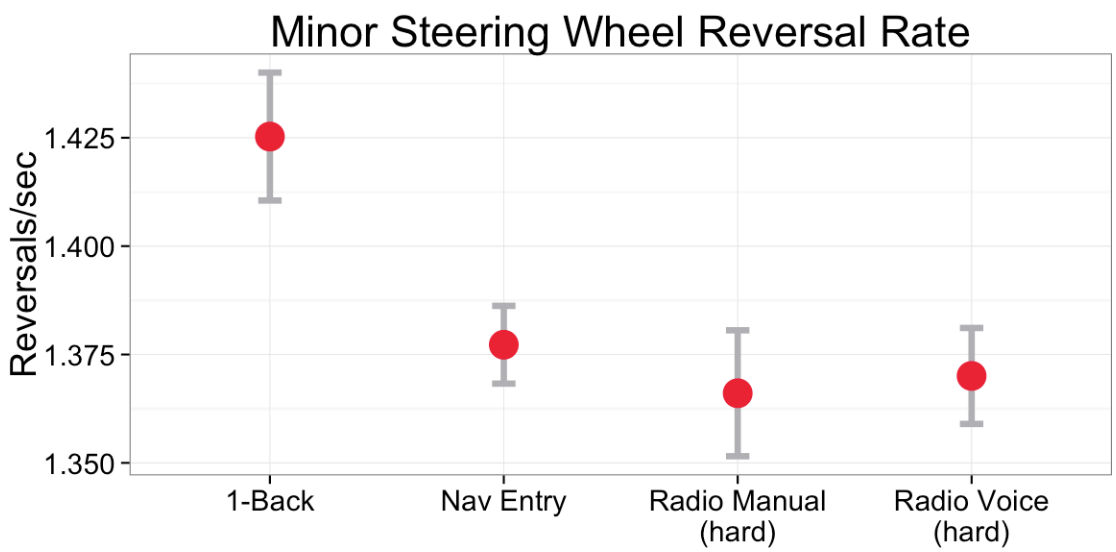


Figure 66: Minor SWR across selected reference tasks.

Minor SWRs did not vary significantly across the selected reference tasks ($X^2 = 6.5$, $p = .090$, Friedman test), though the 1-back task does have a nominally higher minor SWR than the other three tasks. This pattern of results is generally consistent with Study 1.

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 33: Mean and (SD) of minor SWR per second by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	1.16 (0.2)	1.18 (0.2)	1.28 (0.2)	1.32 (0.2)	1.23 (0.2)
Nav Entry	1.33 (0.2)	1.31 (0.2)	1.40 (0.2)	1.46 (0.2)	1.38 (0.2)
Nav Cancel	1.32 (0.3)	1.35 (0.2)	1.44 (0.2)	1.42 (0.2)	1.38 (0.2)
Radio Manual Easy	1.25 (0.3)	1.25 (0.2)	1.31 (0.3)	1.32 (0.2)	1.28 (0.3)
Radio Manual Hard	1.29 (0.2)	1.30 (0.2)	1.45 (0.1)	1.42 (0.2)	1.37 (0.2)
Radio Voice Easy	1.30 (0.2)	1.26 (0.2)	1.41 (0.3)	1.48 (0.2)	1.36 (0.2)
Radio Voice Hard	1.31 (0.2)	1.33 (0.2)	1.38 (0.2)	1.46 (0.2)	1.37 (0.2)
N-Back Reference	1.21 (0.2)	1.23 (0.2)	1.44 (0.4)	1.39 (0.3)	1.32 (0.3)
Blank-Back	1.17 (0.2)	1.20 (0.2)	1.39 (0.3)	1.43 (0.2)	1.30 (0.3)
0-Back	1.29 (0.2)	1.31 (0.2)	1.39 (0.3)	1.43 (0.2)	1.36 (0.2)
1-Back	1.35 (0.3)	1.34 (0.2)	1.44 (0.2)	1.56 (0.3)	1.43 (0.3)
2-Back	1.35 (0.3)	1.31 (0.2)	1.44 (0.2)	1.51 (0.2)	1.41 (0.2)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	1.18 (0.2)	1.24 (0.1)	1.29 (0.2)	1.31 (0.1)	1.26 (0.2)
Nav Entry	1.37 (0.2)	1.37 (0.1)	1.37 (0.2)	1.46 (0.2)	1.39 (0.2)
Nav Cancel	1.32 (0.2)	1.40 (0.2)	1.31 (0.2)	1.42 (0.2)	1.36 (0.2)
Radio Manual Easy	1.39 (0.2)	1.39 (0.2)	1.34 (0.4)	1.34 (0.3)	1.36 (0.3)
Radio Manual Hard	1.30 (0.1)	1.38 (0.2)	1.44 (0.2)	1.45 (0.2)	1.39 (0.2)
Radio Voice Easy	1.27 (0.2)	1.31 (0.1)	1.41 (0.3)	1.47 (0.3)	1.37 (0.3)
Radio Voice Hard	1.25 (0.2)	1.35 (0.1)	1.38 (0.2)	1.46 (0.2)	1.36 (0.2)
N-Back Reference	1.20 (0.1)	1.23 (0.2)	1.36 (0.2)	1.44 (0.2)	1.31 (0.2)
Blank-Back	1.14 (0.2)	1.30 (0.2)	1.38 (0.3)	1.41 (0.3)	1.32 (0.3)
0-Back	1.31 (0.2)	1.30 (0.1)	1.39 (0.3)	1.43 (0.2)	1.36 (0.2)
1-Back	1.34 (0.1)	1.35 (0.1)	1.40 (0.3)	1.58 (0.3)	1.42 (0.2)
2-Back	1.31 (0.3)	1.32 (0.2)	1.50 (0.3)	1.48 (0.2)	1.41 (0.2)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	1.13 (0.2)	1.12 (0.2)	1.26 (0.2)	1.33 (0.2)	1.21 (0.2)
Nav Entry	1.28 (0.2)	1.25 (0.2)	1.44 (0.2)	1.47 (0.3)	1.36 (0.2)
Nav Cancel	1.32 (0.3)	1.30 (0.2)	1.56 (0.2)	1.42 (0.3)	1.40 (0.3)
Radio Manual Easy	1.12 (0.3)	1.11 (0.2)	1.28 (0.2)	1.30 (0.2)	1.20 (0.2)
Radio Manual Hard	1.29 (0.2)	1.22 (0.2)	1.46 (0.1)	1.40 (0.2)	1.34 (0.2)
Radio Voice Easy	1.33 (0.2)	1.20 (0.2)	1.41 (0.1)	1.50 (0.2)	1.36 (0.2)
Radio Voice Hard	1.37 (0.2)	1.31 (0.2)	1.37 (0.2)	1.46 (0.3)	1.38 (0.2)
N-Back Reference	1.21 (0.3)	1.24 (0.2)	1.51 (0.5)	1.34 (0.3)	1.33 (0.3)
Blank-Back	1.20 (0.2)	1.10 (0.2)	1.40 (0.3)	1.45 (0.2)	1.28 (0.2)
0-Back	1.27 (0.2)	1.32 (0.2)	1.40 (0.2)	1.43 (0.3)	1.36 (0.2)
1-Back	1.35 (0.3)	1.33 (0.2)	1.48 (0.3)	1.55 (0.3)	1.43 (0.3)
2-Back	1.39 (0.2)	1.30 (0.2)	1.39 (0.2)	1.54 (0.2)	1.41 (0.2)

Table 34: Mean and (SD) of minor SWR per minute by task, broken down by age group and gender.

Combined	20-24	25-39	40-54	55+	(all)
Baseline	69.38 (10.4)	70.92 (9.4)	76.68 (10.6)	79.21 (9.8)	74.05 (10.6)
Nav Entry	79.77 (10.5)	78.73 (9.4)	84.28 (12.2)	87.77 (13.8)	82.64 (11.9)
Nav Cancel	79.04 (15.3)	80.93 (10.9)	86.14 (14.1)	85.13 (12.5)	82.81 (13.3)
Radio Manual Easy	75.02 (15.5)	74.85 (15.0)	78.56 (19.5)	79.29 (14.4)	76.93 (15.9)
Radio Manual Hard	77.63 (10.1)	78.03 (11.0)	86.79 (8.0)	85.40 (11.3)	81.96 (10.8)
Radio Voice Easy	77.98 (13.8)	75.35 (9.3)	84.59 (15.0)	89.02 (14.1)	81.73 (14.0)
Radio Voice Hard	78.75 (11.3)	79.95 (10.6)	82.56 (11.6)	87.56 (13.9)	82.20 (12.1)
N-Back Reference	72.43 (12.5)	74.00 (12.6)	86.25 (21.2)	83.38 (15.5)	79.23 (16.6)
Blank-Back	70.43 (11.5)	71.75 (13.6)	83.12 (15.1)	85.75 (14.6)	78.00 (15.1)
0-Back	77.29 (12.8)	78.88 (10.4)	83.50 (16.7)	85.50 (13.8)	81.42 (13.7)
1-Back	80.71 (15.3)	80.38 (9.8)	86.50 (14.9)	93.88 (17.4)	85.52 (15.3)
2-Back	81.29 (15.0)	78.88 (11.2)	86.62 (14.3)	90.62 (12.9)	84.45 (13.9)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	70.87 (10.2)	74.67 (7.8)	77.66 (10.4)	78.53 (8.8)	75.43 (9.4)
Nav Entry	82.45 (9.1)	82.35 (7.5)	82.32 (14.0)	87.51 (13.3)	83.66 (11.0)
Nav Cancel	79.13 (10.3)	83.91 (9.6)	78.44 (12.1)	85.18 (9.2)	81.66 (10.3)
Radio Manual Easy	83.12 (11.5)	83.24 (10.7)	80.28 (24.2)	80.67 (16.9)	81.83 (16.0)
Radio Manual Hard	77.78 (8.4)	83.05 (10.5)	86.19 (10.4)	86.74 (10.2)	83.44 (10.1)
Radio Voice Easy	76.34 (13.9)	78.59 (7.6)	84.85 (20.8)	88.21 (16.0)	82.00 (15.4)
Radio Voice Hard	75.30 (10.0)	81.29 (8.0)	82.88 (13.0)	87.58 (13.4)	81.76 (11.6)
N-Back Reference	72.00 (7.4)	73.50 (12.9)	81.75 (12.8)	86.50 (13.8)	78.87 (13.1)
Blank-Back	68.67 (12.6)	77.75 (12.8)	82.50 (16.1)	84.50 (17.8)	79.00 (15.5)
0-Back	78.33 (13.0)	78.25 (8.1)	83.25 (20.5)	85.50 (11.9)	81.53 (13.8)
1-Back	80.33 (7.2)	80.75 (8.1)	84.25 (15.1)	95.00 (18.8)	85.40 (14.2)
2-Back	78.67 (17.2)	79.50 (9.1)	89.75 (15.4)	89.00 (14.4)	84.60 (14.3)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	67.88 (11.0)	67.17 (9.8)	75.70 (11.4)	79.89 (11.3)	72.66 (11.7)
Nav Entry	77.08 (11.6)	75.10 (10.2)	86.23 (10.5)	88.03 (15.2)	81.61 (12.8)
Nav Cancel	78.94 (19.9)	77.95 (12.0)	93.84 (12.1)	85.09 (15.7)	83.95 (15.9)
Radio Manual Easy	66.93 (15.2)	66.45 (14.3)	76.84 (14.8)	77.90 (12.4)	72.03 (14.6)
Radio Manual Hard	77.49 (12.2)	73.01 (9.5)	87.38 (5.4)	84.06 (12.8)	80.49 (11.4)
Radio Voice Easy	79.62 (14.4)	72.11 (10.2)	84.33 (7.0)	89.82 (12.9)	81.47 (12.8)
Radio Voice Hard	82.21 (12.1)	78.60 (13.2)	82.23 (10.8)	87.54 (15.3)	82.64 (12.7)
N-Back Reference	72.75 (15.8)	74.50 (13.2)	90.75 (27.4)	80.25 (17.3)	79.56 (19.6)
Blank-Back	71.75 (11.3)	65.75 (12.3)	83.75 (15.2)	87.00 (11.8)	77.06 (15.0)
0-Back	76.50 (13.4)	79.50 (12.9)	83.75 (13.2)	85.50 (16.3)	81.31 (13.8)
1-Back	81.00 (20.0)	80.00 (11.8)	88.75 (15.2)	92.75 (17.2)	85.62 (16.4)
2-Back	83.25 (14.1)	78.25 (13.7)	83.50 (13.3)	92.25 (12.0)	84.31 (13.6)

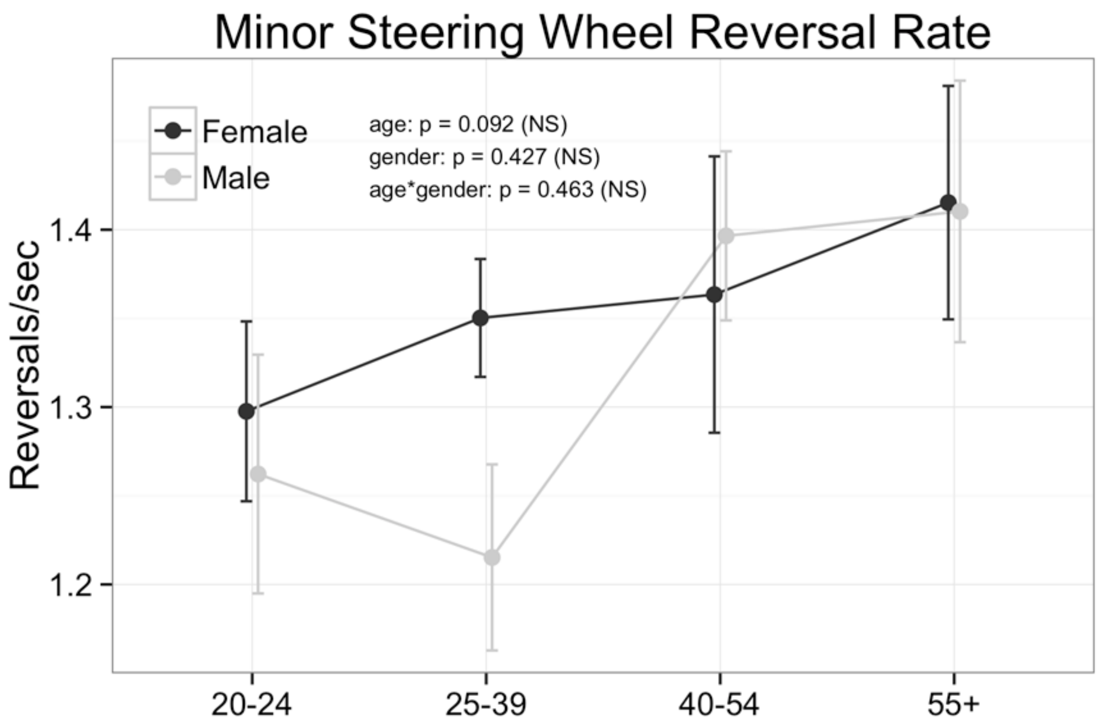


Figure 67: Minor SWR across age groups and genders.

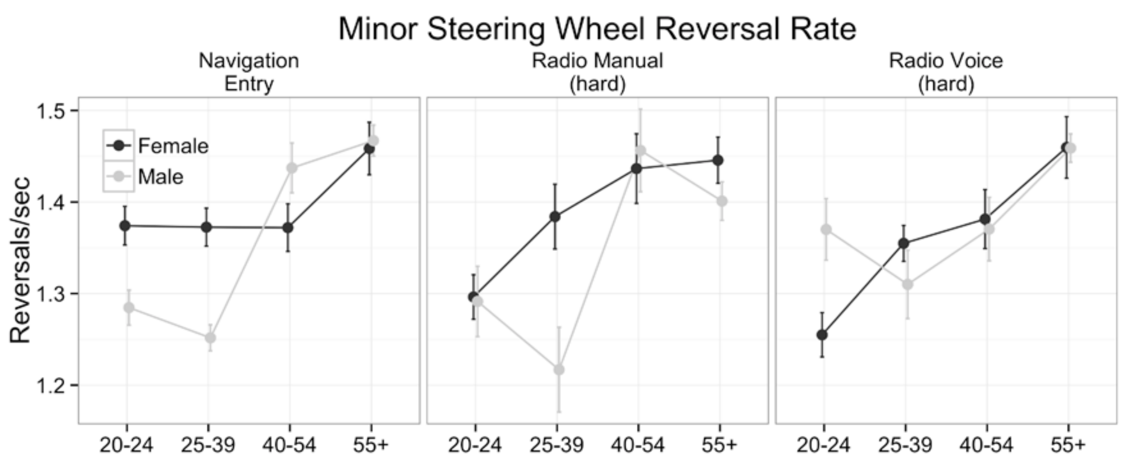


Figure 68: Minor SWR by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing minor SWRs by age and gender, there are no significant effects for either factor, though minor SWR does tend to increase with age.

Major Steering Wheel Reversal Rate

Major (large) steering wheel reversals are similar to the minor reversal metric outlined above. They are counted and classified with a larger gap size of 3º, and indicate grosser adjustments in lateral vehicle control.

Table 35: Mean (and standard deviation) of major SWR per second by training type.

	Structured Training	Self-Guided Training	(combined)
Baseline	0.12 (0.0)	0.10 (0.0)	0.11 (0.0)
Nav Entry	0.13 (0.1)	0.12 (0.0)	0.13 (0.1)
Nav Cancel	0.13 (0.1)	0.12 (0.1)	0.12 (0.1)
Radio Manual Easy	0.09 (0.1)	0.10 (0.1)	0.09 (0.1)
Radio Manual Hard	0.16 (0.1)	0.16 (0.1)	0.16 (0.1)
Radio Voice Easy	0.11 (0.1)	0.12 (0.1)	0.11 (0.1)
Radio Voice Hard	0.11 (0.1)	0.12 (0.1)	0.11 (0.1)

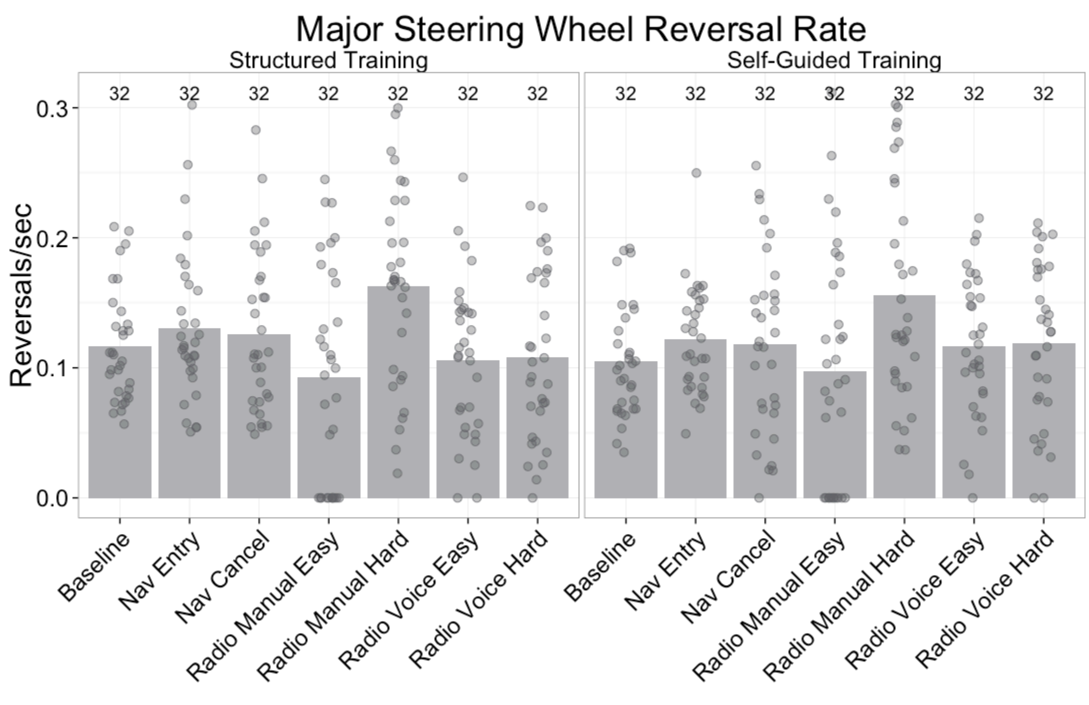


Figure 69: Major SWR with bars representing means and solid circles individual participants.

Major SWR were not affected by training condition (p = .902).

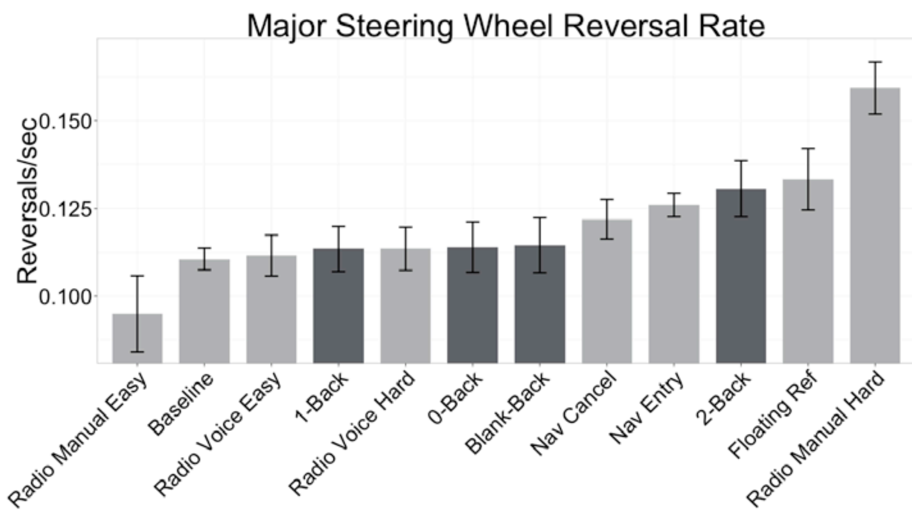


Figure 70: Major SWR means and adjusted standard errors. Levels of the n-back task are shaded in dark gray.

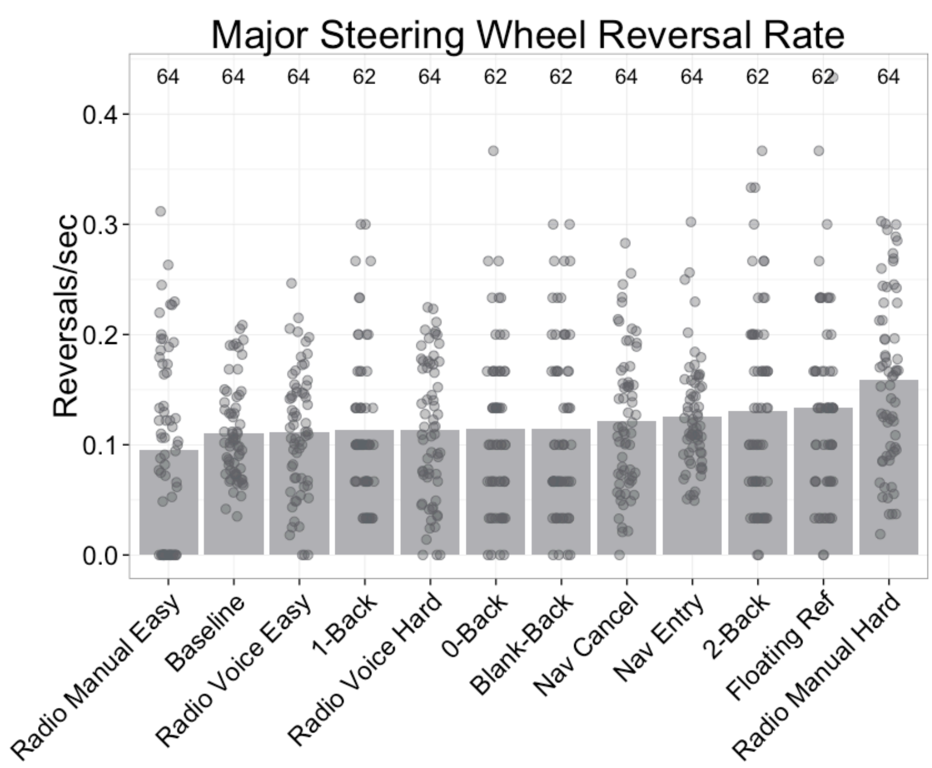


Figure 71: Major SWR by task and training condition.

Consistent with Study 1, major SWRs were highest during the Radio Manual Hard task.

In this study, the Radio Manual Easy task showed the nominally lowest major SWR rate, while it showed the second highest mean value in Study 1. Noting that the ranking pattern for the minor SWR for the Radio Manual Easy task also showed divergence between Study 1 and 2, it may be that the relatively discrete nature of a single button press task may significantly impact the reliability of rate based measures such as these.

Statistical Comparison of Selected Tasks

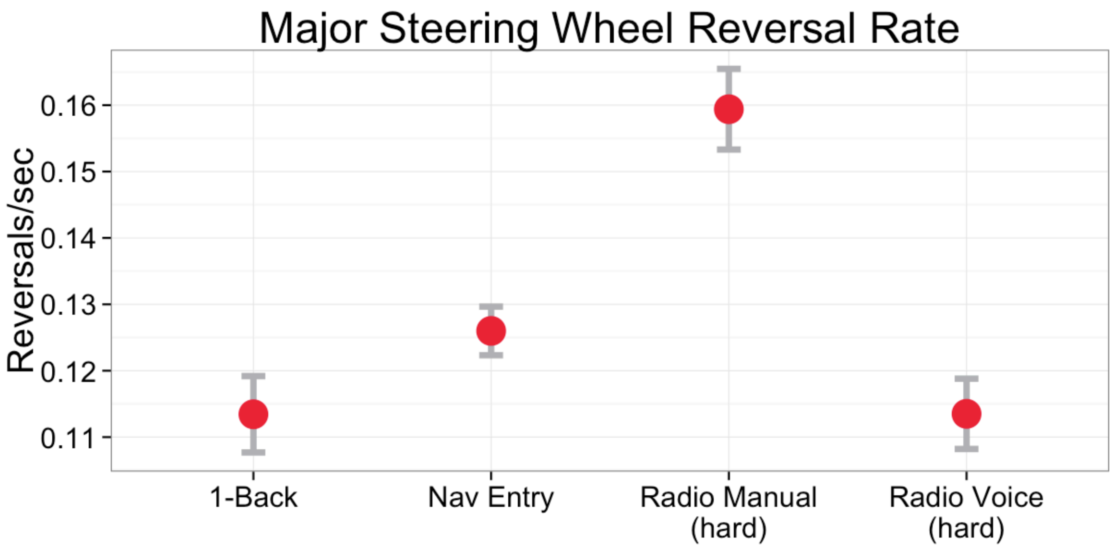


Figure 72: Major SWR across selected reference tasks.

Major SWR rate varied significantly across the four reference tasks ($X^2 = 23.7$, $p < .001$, Friedman test). In contrast to the pattern observed with minor SWR, in which the 1-back task had a nominally higher SWR rate compared to the other tasks, the effect for major SWR is driven by the higher rates observed during the Radio Manual Hard task ($p < .001$ in individual comparisons to the other reference tasks, Wilcoxon signed rank test). This pattern of results is consistent with those of Study 1.Major SWR rates are often interpreted as reflecting increased steering effort to correct lane keeping errors.

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 36: Mean and (SD) of major SWR per second by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	0.11 (0.0)	0.12 (0.0)	0.11 (0.0)	0.10 (0.1)	0.11 (0.0)
Nav Entry	0.12 (0.0)	0.13 (0.0)	0.14 (0.1)	0.12 (0.0)	0.13 (0.1)
Nav Cancel	0.13 (0.1)	0.13 (0.1)	0.11 (0.1)	0.11 (0.1)	0.12 (0.1)
Radio Manual Easy	0.09 (0.1)	0.11 (0.1)	0.07 (0.1)	0.11 (0.1)	0.09 (0.1)
Radio Manual Hard	0.15 (0.1)	0.17 (0.1)	0.18 (0.1)	0.14 (0.1)	0.16 (0.1)
Radio Voice Easy	0.11 (0.0)	0.11 (0.1)	0.11 (0.1)	0.12 (0.1)	0.11 (0.1)
Radio Voice Hard	0.11 (0.1)	0.11 (0.1)	0.12 (0.1)	0.11 (0.1)	0.11 (0.1)
N-Back Reference	0.14 (0.1)	0.13 (0.1)	0.16 (0.1)	0.11 (0.1)	0.13 (0.1)
Blank-Back	0.10 (0.1)	0.14 (0.1)	0.11 (0.1)	0.11 (0.1)	0.11 (0.1)
0-Back	0.12 (0.1)	0.12 (0.1)	0.11 (0.1)	0.11 (0.1)	0.11 (0.1)
1-Back	0.10 (0.0)	0.14 (0.1)	0.12 (0.1)	0.10 (0.1)	0.11 (0.1)
2-Back	0.11 (0.1)	0.15 (0.1)	0.12 (0.1)	0.13 (0.1)	0.13 (0.1)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	0.12 (0.0)	0.11 (0.0)	0.12 (0.0)	0.11 (0.1)	0.11 (0.0)
Nav Entry	0.12 (0.0)	0.11 (0.0)	0.16 (0.1)	0.11 (0.0)	0.12 (0.0)
Nav Cancel	0.15 (0.1)	0.13 (0.1)	0.13 (0.1)	0.12 (0.1)	0.13 (0.1)
Radio Manual Easy	0.13 (0.1)	0.13 (0.1)	0.09 (0.1)	0.12 (0.1)	0.11 (0.1)
Radio Manual Hard	0.14 (0.1)	0.16 (0.1)	0.17 (0.1)	0.13 (0.1)	0.15 (0.1)
Radio Voice Easy	0.11 (0.0)	0.09 (0.1)	0.10 (0.1)	0.11 (0.1)	0.10 (0.1)
Radio Voice Hard	0.09 (0.1)	0.09 (0.1)	0.10 (0.1)	0.10 (0.1)	0.10 (0.1)
N-Back Reference	0.13 (0.1)	0.10 (0.1)	0.16 (0.1)	0.11 (0.1)	0.12 (0.1)
Blank-Back	0.09 (0.1)	0.14 (0.1)	0.12 (0.1)	0.09 (0.1)	0.11 (0.1)
0-Back	0.09 (0.1)	0.09 (0.1)	0.12 (0.1)	0.11 (0.1)	0.10 (0.1)
1-Back	0.10 (0.1)	0.13 (0.1)	0.13 (0.1)	0.07 (0.0)	0.11 (0.1)
2-Back	0.13 (0.1)	0.16 (0.1)	0.13 (0.1)	0.12 (0.1)	0.14 (0.1)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	0.10 (0.0)	0.12 (0.0)	0.11 (0.0)	0.10 (0.0)	0.11 (0.0)
Nav Entry	0.11 (0.0)	0.14 (0.1)	0.13 (0.1)	0.13 (0.0)	0.13 (0.1)
Nav Cancel	0.12 (0.1)	0.13 (0.1)	0.09 (0.1)	0.11 (0.1)	0.11 (0.1)
Radio Manual Easy	0.06 (0.1)	0.09 (0.1)	0.05 (0.0)	0.09 (0.1)	0.08 (0.1)
Radio Manual Hard	0.16 (0.1)	0.17 (0.1)	0.19 (0.1)	0.15 (0.1)	0.17 (0.1)
Radio Voice Easy	0.11 (0.0)	0.14 (0.1)	0.12 (0.1)	0.12 (0.0)	0.12 (0.1)
Radio Voice Hard	0.13 (0.1)	0.13 (0.1)	0.14 (0.0)	0.12 (0.0)	0.13 (0.1)
N-Back Reference	0.15 (0.0)	0.15 (0.1)	0.16 (0.1)	0.10 (0.1)	0.14 (0.1)
Blank-Back	0.11 (0.1)	0.14 (0.1)	0.10 (0.1)	0.13 (0.1)	0.12 (0.1)
0-Back	0.14 (0.1)	0.15 (0.1)	0.09 (0.1)	0.12 (0.0)	0.12 (0.1)
1-Back	0.10 (0.0)	0.14 (0.1)	0.11 (0.1)	0.12 (0.1)	0.12 (0.1)
2-Back	0.10 (0.1)	0.15 (0.1)	0.12 (0.1)	0.13 (0.1)	0.12 (0.1)

Table 37: Mean and (SD) of major SWR per minute by task, broken down by age group and gender.

Combined	20-24	25-39	40-54	55+	(all)
Baseline	6.68 (1.7)	7.06 (2.7)	6.73 (2.7)	6.06 (3.2)	6.63 (2.6)
Nav Entry	6.96 (2.6)	7.57 (2.9)	8.50 (3.7)	7.20 (2.8)	7.56 (3.0)
Nav Cancel	7.99 (4.1)	7.98 (3.2)	6.61 (4.5)	6.69 (3.8)	7.32 (3.9)
Radio Manual Easy	5.68 (6.1)	6.55 (5.3)	4.22 (5.0)	6.32 (4.3)	5.70 (5.2)
Radio Manual Hard	8.84 (4.1)	9.97 (5.7)	10.79 (4.3)	8.65 (4.7)	9.56 (4.7)
Radio Voice Easy	6.38 (2.3)	6.79 (4.1)	6.64 (4.2)	6.96 (3.0)	6.69 (3.4)
Radio Voice Hard	6.69 (4.1)	6.67 (4.1)	7.33 (3.9)	6.55 (3.4)	6.81 (3.8)
N-Back Reference	8.43 (5.2)	7.62 (3.7)	9.50 (6.6)	6.50 (4.6)	8.00 (5.1)
Blank-Back	6.00 (4.9)	8.38 (4.9)	6.50 (4.6)	6.50 (5.2)	6.87 (4.9)
0-Back	7.14 (4.0)	7.00 (5.4)	6.38 (4.7)	6.88 (4.1)	6.84 (4.5)
1-Back	5.86 (2.4)	8.25 (4.4)	7.25 (5.2)	5.75 (3.6)	6.81 (4.1)
2-Back	6.86 (5.6)	9.12 (5.4)	7.50 (5.0)	7.75 (5.9)	7.84 (5.4)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	7.11 (2.3)	6.76 (2.7)	7.02 (2.8)	6.42 (3.8)	6.83 (2.8)
Nav Entry	7.34 (2.9)	6.52 (2.6)	9.30 (3.0)	6.71 (2.8)	7.47 (2.9)
Nav Cancel	8.96 (4.4)	7.92 (3.7)	7.82 (3.1)	6.96 (4.0)	7.91 (3.7)
Radio Manual Easy	7.68 (6.6)	7.56 (6.3)	5.19 (6.5)	6.97 (4.3)	6.85 (5.8)
Radio Manual Hard	8.27 (4.0)	9.50 (5.4)	10.12 (4.6)	8.04 (5.7)	8.98 (4.8)
Radio Voice Easy	6.44 (1.9)	5.34 (4.3)	6.06 (3.4)	6.78 (3.4)	6.16 (3.2)
Radio Voice Hard	5.48 (4.3)	5.57 (3.9)	6.21 (4.8)	6.17 (3.9)	5.86 (4.1)
N-Back Reference	7.67 (7.5)	6.00 (3.4)	9.50 (6.3)	6.75 (5.8)	7.47 (5.7)
Blank-Back	5.33 (4.8)	8.50 (5.1)	7.25 (5.5)	5.25 (5.2)	6.67 (5.1)
0-Back	5.67 (4.3)	5.25 (4.1)	7.25 (5.4)	6.50 (5.4)	6.20 (4.7)
1-Back	6.00 (3.3)	8.00 (4.5)	8.00 (5.6)	4.25 (2.3)	6.60 (4.3)
2-Back	8.00 (7.5)	9.50 (4.8)	7.75 (5.5)	7.50 (6.7)	8.20 (5.8)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Baseline	6.24 (0.8)	7.37 (2.9)	6.44 (2.7)	5.71 (2.8)	6.44 (2.4)
Nav Entry	6.58 (2.5)	8.62 (3.0)	7.70 (4.3)	7.70 (2.9)	7.65 (3.2)
Nav Cancel	7.01 (3.8)	8.03 (3.0)	5.41 (5.5)	6.41 (4.0)	6.72 (4.1)
Radio Manual Easy	3.68 (5.3)	5.55 (4.2)	3.26 (2.9)	5.67 (4.5)	4.54 (4.2)
Radio Manual Hard	9.42 (4.3)	10.45 (6.3)	11.46 (4.2)	9.26 (3.8)	10.15 (4.6)
Radio Voice Easy	6.32 (2.9)	8.23 (3.6)	7.23 (4.9)	7.14 (2.9)	7.23 (3.6)
Radio Voice Hard	7.89 (3.8)	7.78 (4.3)	8.45 (2.6)	6.93 (3.0)	7.76 (3.3)
N-Back Reference	9.00 (2.8)	9.25 (3.5)	9.50 (7.2)	6.25 (3.5)	8.50 (4.6)
Blank-Back	6.50 (5.2)	8.25 (4.9)	5.75 (3.6)	7.75 (5.3)	7.06 (4.7)
0-Back	8.25 (3.6)	8.75 (6.2)	5.50 (4.1)	7.25 (2.6)	7.44 (4.3)
1-Back	5.75 (1.7)	8.50 (4.5)	6.50 (5.1)	7.25 (4.3)	7.00 (4.0)
2-Back	6.00 (4.1)	8.75 (6.2)	7.25 (4.8)	8.00 (5.5)	7.50 (5.1)

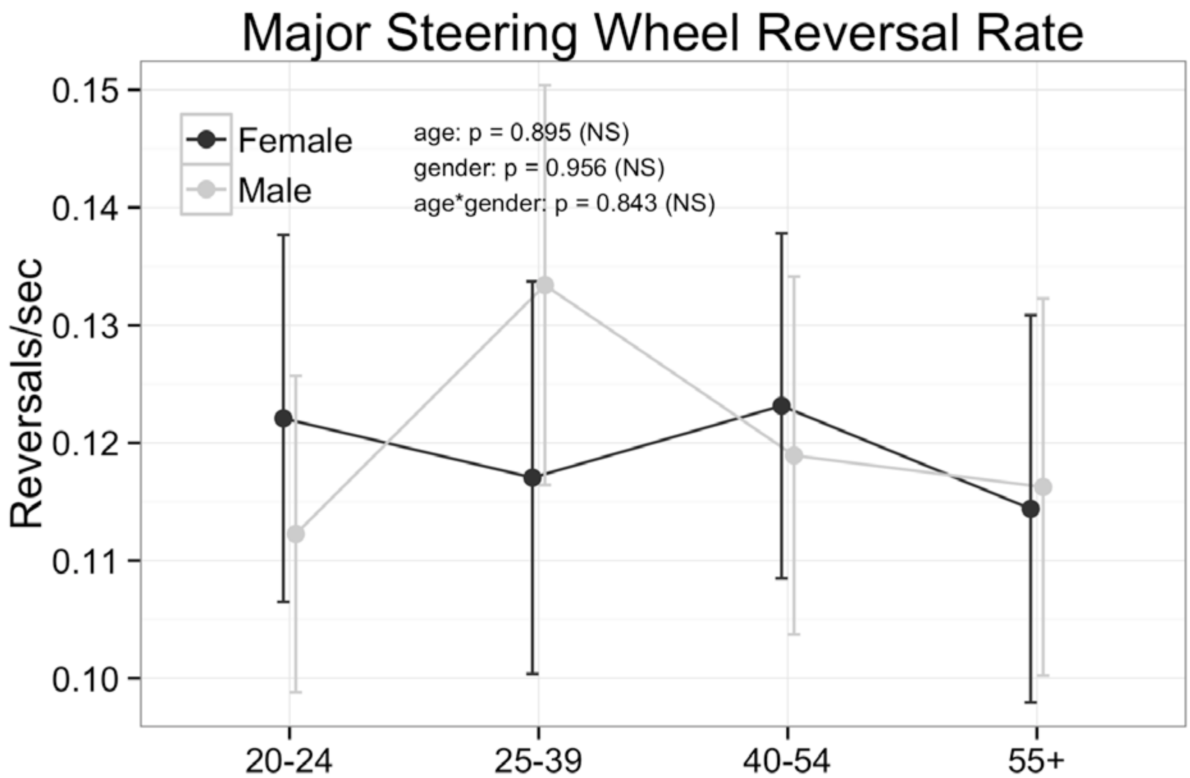


Figure 73: Major SWR across age groups and genders.

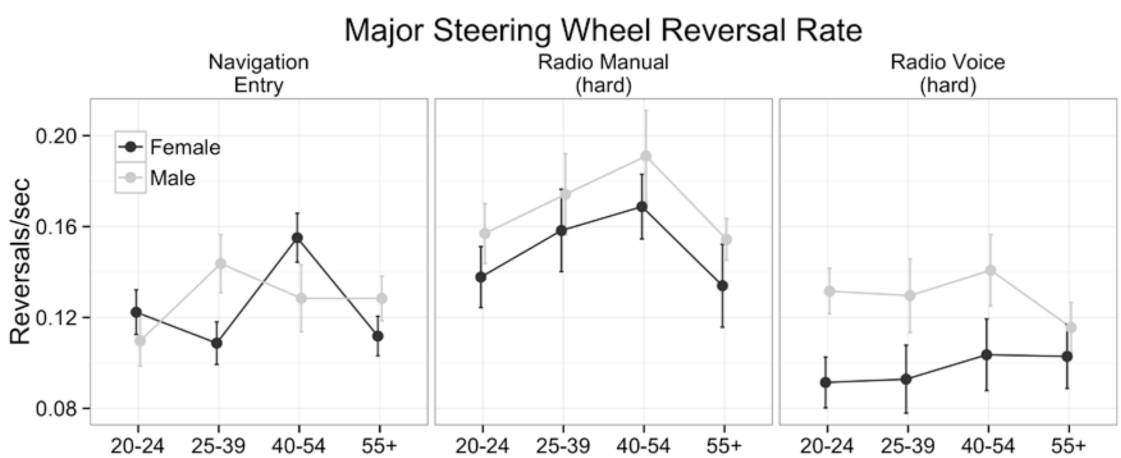


Figure 74: Major SWR by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing major SWR by age and gender, we see no significant main effects. Subplots for the Radio Manual Hard and Radio Voice Hard suggest possible gender trends for these particular tasks, with males tending to show larger SWRs in these cases.

Glance Analyses

NHTSA specifies a minimum sample size of 24 participants (with specified age, gender, and experience characteristics) and mandates that at least 21 out of the 24 participants meet each of the following criteria while performing the "testable task":

- **Percentage of Long Single Duration Glances**. No more than 15 percent (rounded up) of the total number of eye glances away from the forward road scene are of duration greater than 2.0 seconds.
- **Mean Off-Road Single Glance Duration**. The mean duration of all eye glances away from the forward road scene is less than or equal to 2.0 seconds.
- **Total Off-Road Glance Time**. The sum of the durations of each individual participant's eye glances away from the forward road scene is less than or equal to 12.0 seconds.

For samples larger than 24, the same proportional relationship is to be applied such that 85% (rounded up) or more of the participants meet the criteria. NHTSA defines "off-road" as any glance off of the forward roadway, which results in glances to the rear- and side-view mirrors as being included in the "off-road" values. More traditional glance to device metrics are also presented in several of the subsections here.

NHTSA Glance Metrics Summary Table

Table 38 shows the percentage of participants in this study who would meet each of the off-the-forward-roadway glance criteria. Entries for situations where less than 85% of a group meet a threshold are bolded and shown in red.

Table 38: Percentage of participants who would pass each of the three NHTSA glance criteria, by task. Tasks that do not meet the 85% threshold are highlighted in red.

	Mean Single Glance Duration	% of Glances > 2s	Total Glance Time
Nav Entry	100.00%	100.00%	**7.80%**
Nav Cancel	100.00%	100.00%	**59.40%**
Radio Manual Easy	100.00%	96.90%	98.40%
Radio Manual Hard	100.00%	**84.40%**	**46.90%**
Radio Voice Easy	100.00%	100.00%	90.60%
Radio Voice Hard	100.00%	100.00%	95.30%

Mean Single Glance Duration

A driver's mean single glance duration summarizes the average length of glances off the forward roadway. Glances to the rear and side-view mirrors are also considered "off-road" in this metric.

Table 39: Mean (and standard deviation) of mean single glance duration by training type.

	Structured Training	Self-Guided Training	(combined)
Nav Entry	0.82 (0.1)	0.82 (0.2)	0.82 (0.2)
Nav Cancel	0.77 (0.2)	0.79 (0.2)	0.78 (0.2)
Radio Manual Easy	0.87 (0.2)	0.78 (0.3)	0.83 (0.3)
Radio Manual Hard	1.06 (0.3)	1.00 (0.3)	1.03 (0.3)
Radio Voice Easy	0.75 (0.2)	0.67 (0.3)	0.71 (0.2)
Radio Voice Hard	0.72 (0.2)	0.69 (0.2)	0.71 (0.2)

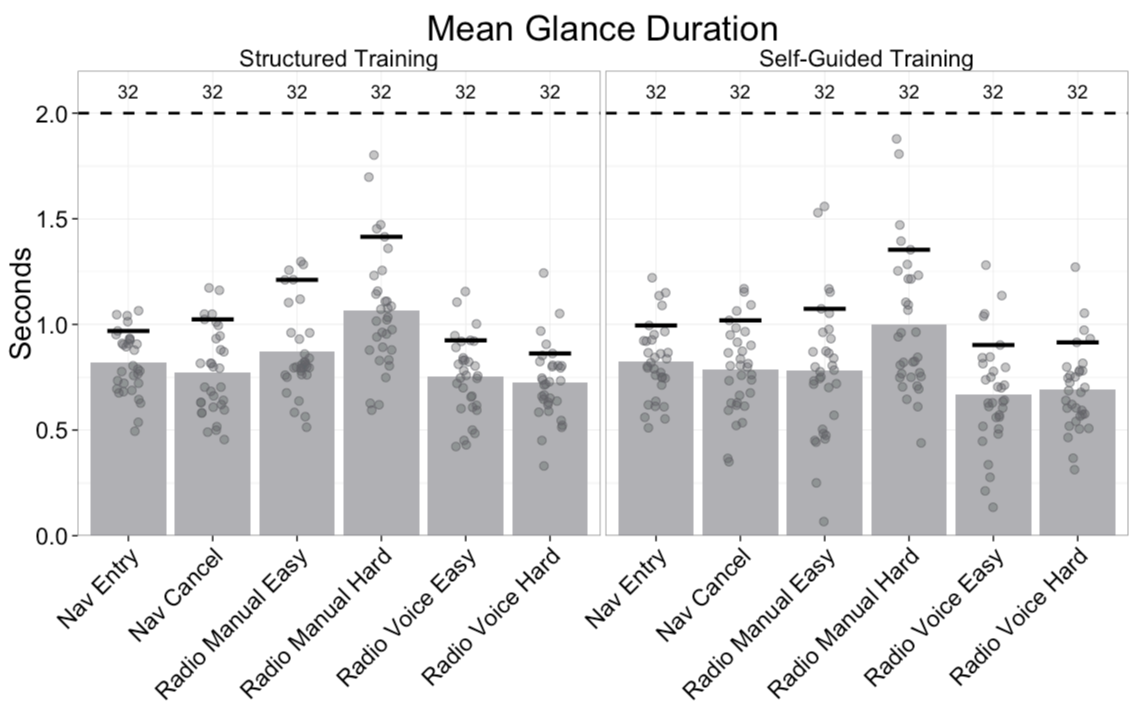


Figure 75: Mean single glance duration by task and training condition.

Mean single glance duration was not affected by training condition (p = .245). Note the relative consistency of the mean glance duration pattern by task across the two training conditions.

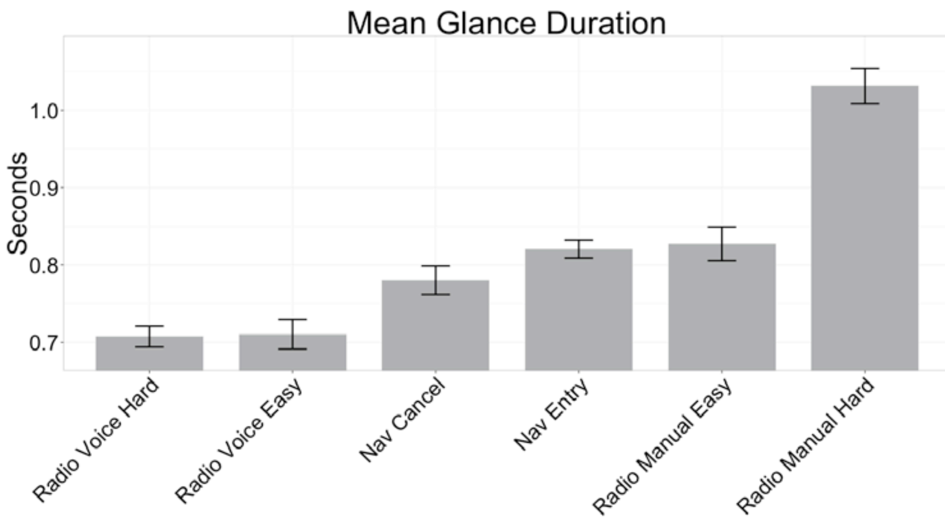


Figure 76: Mean single glance duration collapsed across training conditions showing means and adjusted standard errors.

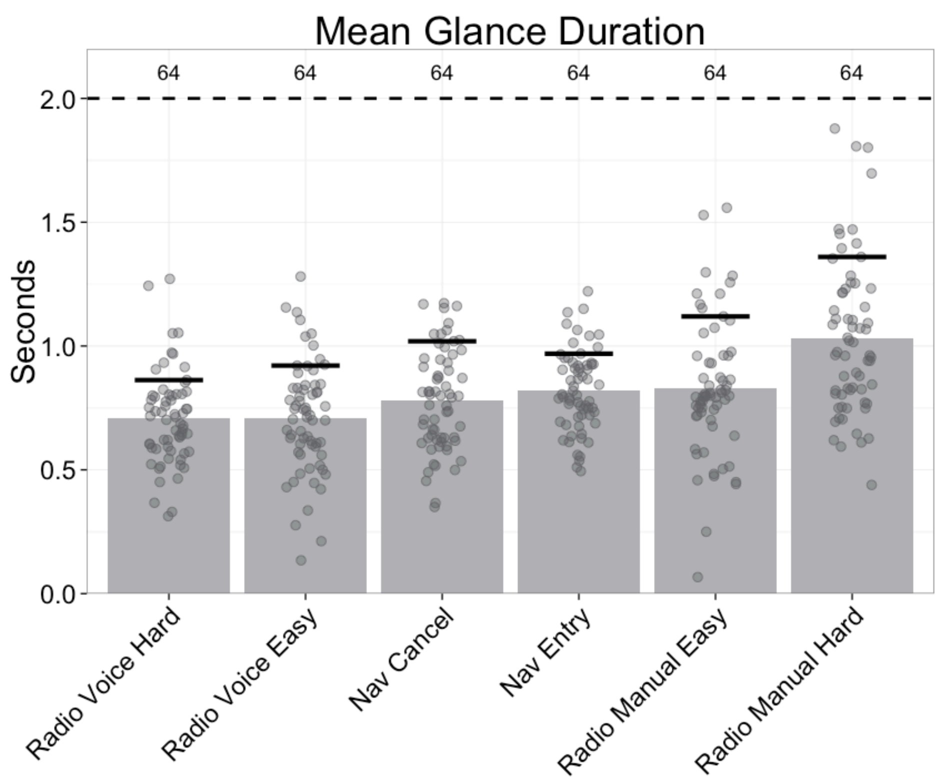


Figure 77: Mean single glance duration with bars representing means and solid circles individual participants.

The Radio Voice Easy and Hard tasks had the shortest glance durations (0.71s for both), while the Radio Manual Hard task had the longest mean glance duration (1.03s). This is broadly consistent with the results of Study 1, in which the Radio Manual Hard task had a mean glance duration of 0.99s, and the Radio Voice Easy and Hard tasks had durations of 0.76s and 0.77s, respectively. Every subject in every task maintained an average glance duration of less than the 2s NHTSA criterion (Figure 78).

Comparison to Study 1

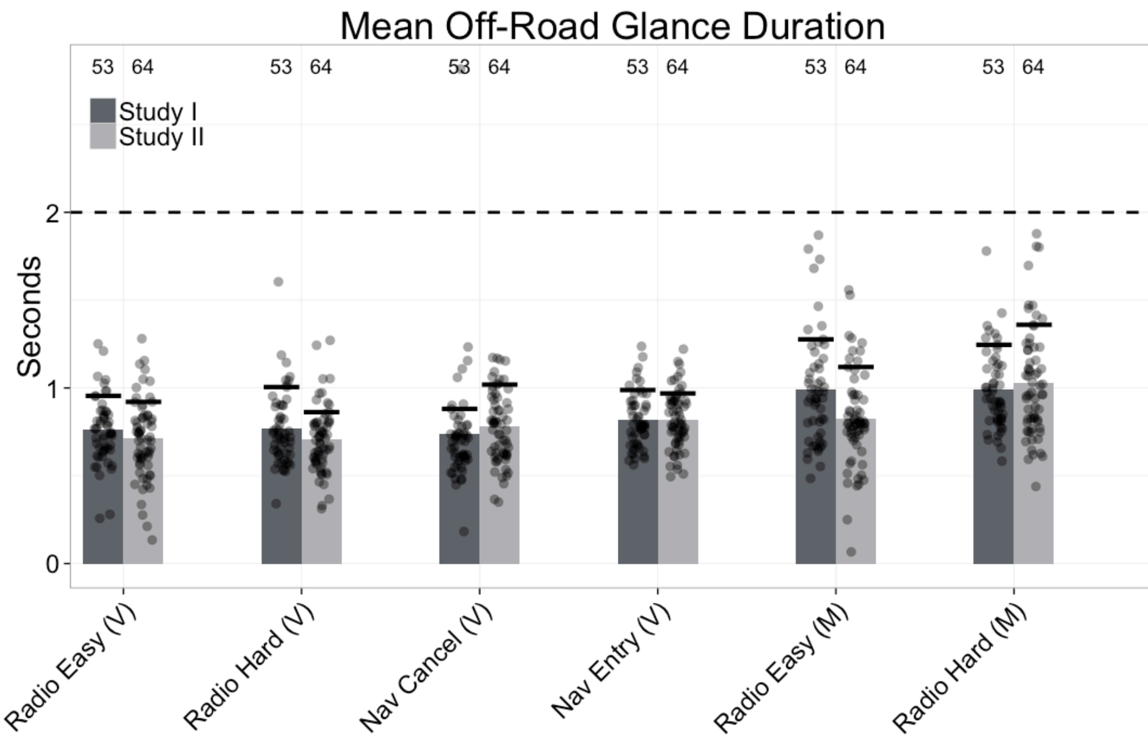


Figure 78: Comparison of mean single glance duration in Study 1 and Study 2.

Mean off-road glance duration is consistently well below the 2 second criterion across all tasks in both studies.

Statistical Comparison of Selected Tasks

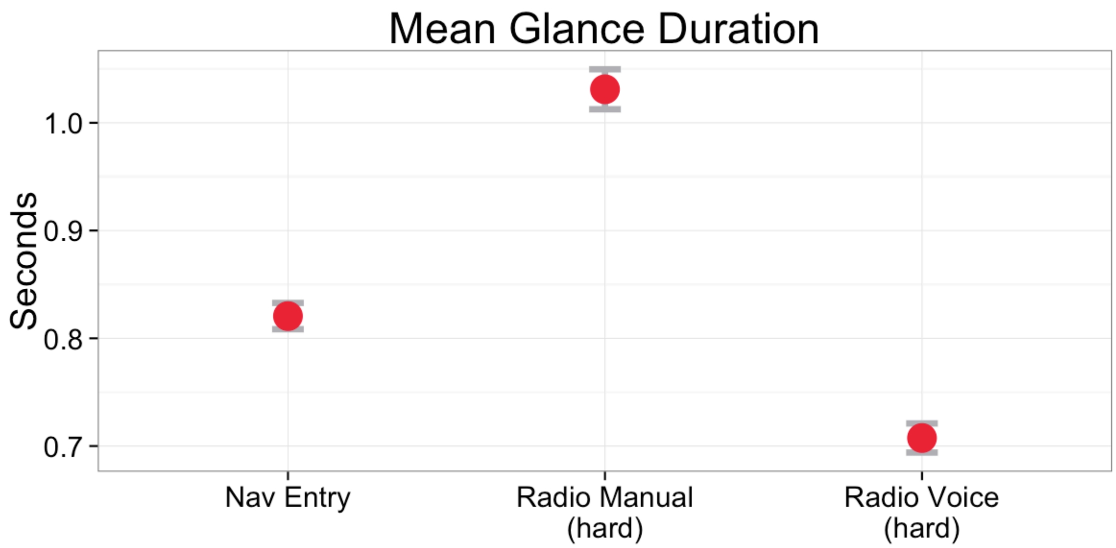


Figure 79: Mean single glance duration across selected reference tasks.

While the 2s criterion may not effectively distinguish between different tasks, mean glance duration itself is significantly different across the three selected reference tasks (X2 = 86.2, $p < .001$, Friedman test). All three tasks were significantly different from one another in post-hoc testing ($p < .001$ for all comparisons, Wilcoxon signed rank tests). The 1-back is not plotted, as no secondary visual demand was present during this task.

The Radio Manual Hard task has the highest mean glance duration, most likely because it typically involves a small number of longer duration glances (see *Percentage of Single Long Duration (> 2s) Glances (TEORT))*. Conversely, the Radio Voice task has the shortest mean glance duration, perhaps because it is a relatively simple task with no direct visual demands beyond possible glance(s) to the push-to-talk voice system button. The Navigation Entry task, a lengthy task that presents a large number of feedback screens, has intermediate mean glance duration.

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 40: Mean and (SD) of mean single glance duration by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	0.77 (0.1)	0.80 (0.2)	0.85 (0.2)	0.86 (0.2)	0.82 (0.2)
Nav Cancel	0.70 (0.1)	0.76 (0.2)	0.82 (0.2)	0.85 (0.2)	0.78 (0.2)
Radio Manual Easy	0.75 (0.3)	0.78 (0.2)	0.93 (0.3)	0.85 (0.3)	0.83 (0.3)
Radio Manual Hard	0.86 (0.2)	1.00 (0.3)	1.20 (0.3)	1.05 (0.3)	1.03 (0.3)
Radio Voice Easy	0.61 (0.2)	0.69 (0.1)	0.77 (0.3)	0.78 (0.2)	0.71 (0.2)
Radio Voice Hard	0.69 (0.2)	0.68 (0.1)	0.72 (0.3)	0.74 (0.2)	0.71 (0.2)
Female	20-24	25-39	40-54	55+	(all)
Nav Entry	0.73 (0.1)	0.79 (0.2)	0.75 (0.1)	0.75 (0.1)	0.75 (0.1)
Nav Cancel	0.73 (0.1)	0.81 (0.3)	0.80 (0.2)	0.75 (0.1)	0.77 (0.2)
Radio Manual Easy	0.77 (0.2)	0.78 (0.2)	0.80 (0.2)	0.71 (0.3)	0.76 (0.2)
Radio Manual Hard	0.88 (0.2)	0.98 (0.2)	0.97 (0.2)	0.84 (0.2)	0.92 (0.2)
Radio Voice Easy	0.59 (0.2)	0.73 (0.2)	0.64 (0.2)	0.66 (0.2)	0.65 (0.2)
Radio Voice Hard	0.70 (0.1)	0.65 (0.1)	0.60 (0.2)	0.62 (0.1)	0.64 (0.1)
Male	20-24	25-39	40-54	55+	(all)
Nav Entry	0.81 (0.2)	0.81 (0.2)	0.96 (0.2)	0.97 (0.1)	0.89 (0.2)
Nav Cancel	0.67 (0.2)	0.71 (0.2)	0.83 (0.2)	0.95 (0.2)	0.79 (0.2)
Radio Manual Easy	0.73 (0.4)	0.78 (0.1)	1.05 (0.4)	0.99 (0.2)	0.89 (0.3)
Radio Manual Hard	0.85 (0.3)	1.03 (0.4)	1.44 (0.3)	1.27 (0.2)	1.14 (0.4)
Radio Voice Easy	0.63 (0.2)	0.64 (0.1)	0.90 (0.3)	0.90 (0.2)	0.77 (0.2)
Radio Voice Hard	0.68 (0.2)	0.72 (0.2)	0.84 (0.3)	0.85 (0.1)	0.77 (0.2)

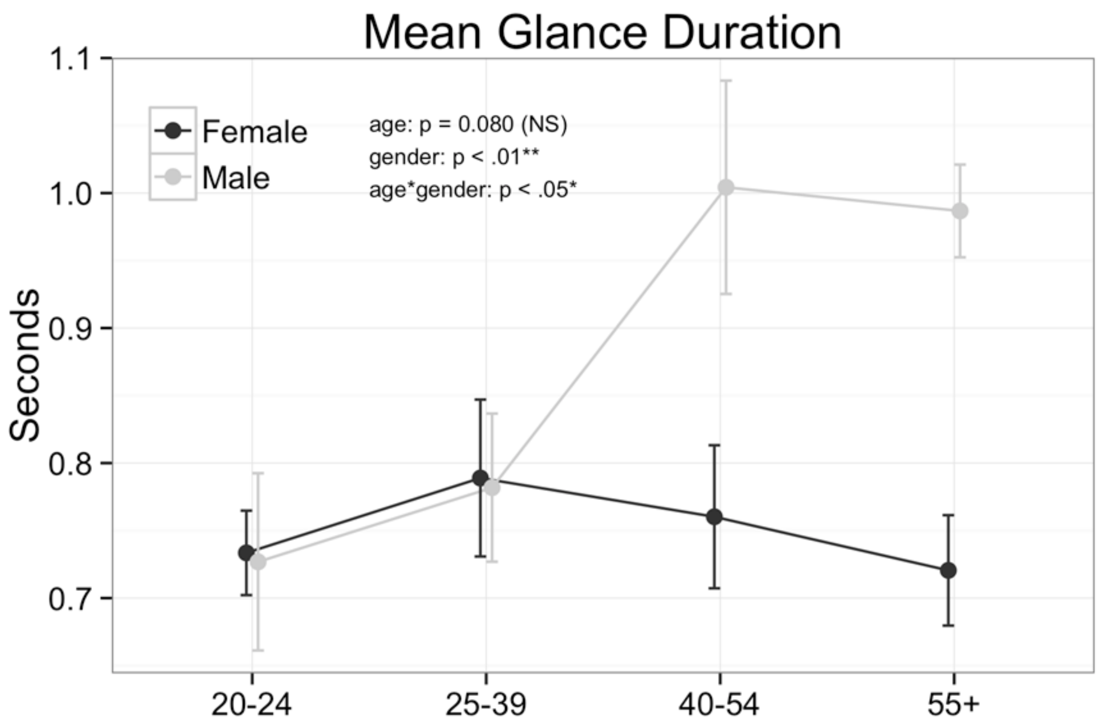


Figure 80: Mean single glance duration across age groups and genders.

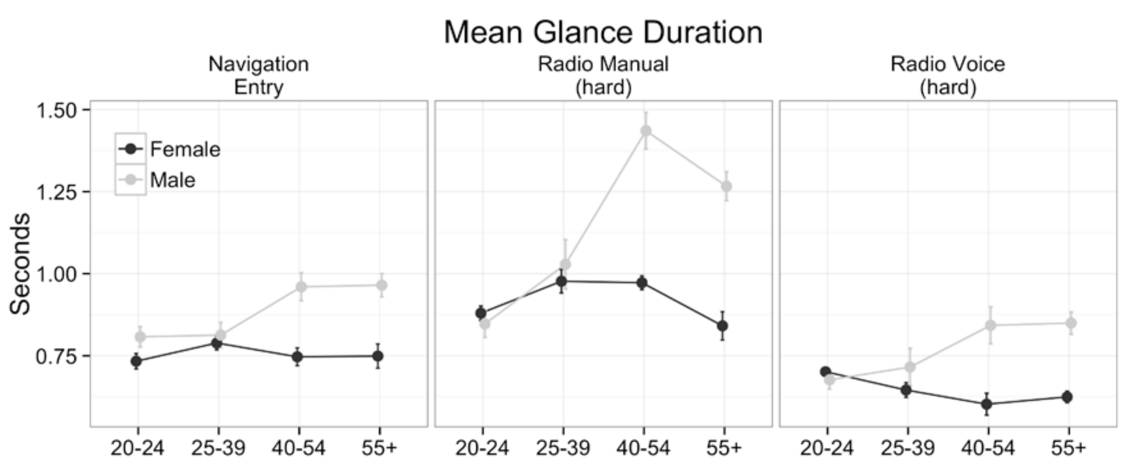


Figure 81: Mean single glance duration by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing mean glance duration by age and gender, we see a significant main effect of gender, as well as an interaction between age and gender. It appears that mean off-road glance duration increases after age 40 among men, while glance durations remain relatively consistent for women.

Percentage of Single Long Duration (> 2s) Glances (TEORT)

A long glance is defined as a single glance of a duration greater than 2s. NHTSA's guidelines on visual/manual distraction (2013) state that for any given task, the percentage of long single glances should be less than 15% per individual.

Table 41: Mean (and standard deviation) of percentage of single glances longer than 2 seconds by training type.

	Structured Training	Self-Guided Training	(combined)
Nav Entry	1.31 (2.5)	1.24 (2.2)	1.27 (2.3)
Nav Cancel	1.23 (2.7)	0.95 (2.1)	1.09 (2.4)
Radio Manual Easy	1.56 (6.2)	1.19 (5.1)	1.38 (5.6)
Radio Manual Hard	6.21 (10.1)	5.72 (9.8)	5.97 (9.9)
Radio Voice Easy	0.94 (3.2)	0.66 (2.5)	0.80 (2.8)
Radio Voice Hard	0.84 (2.4)	0.43 (1.7)	0.64 (2.1)

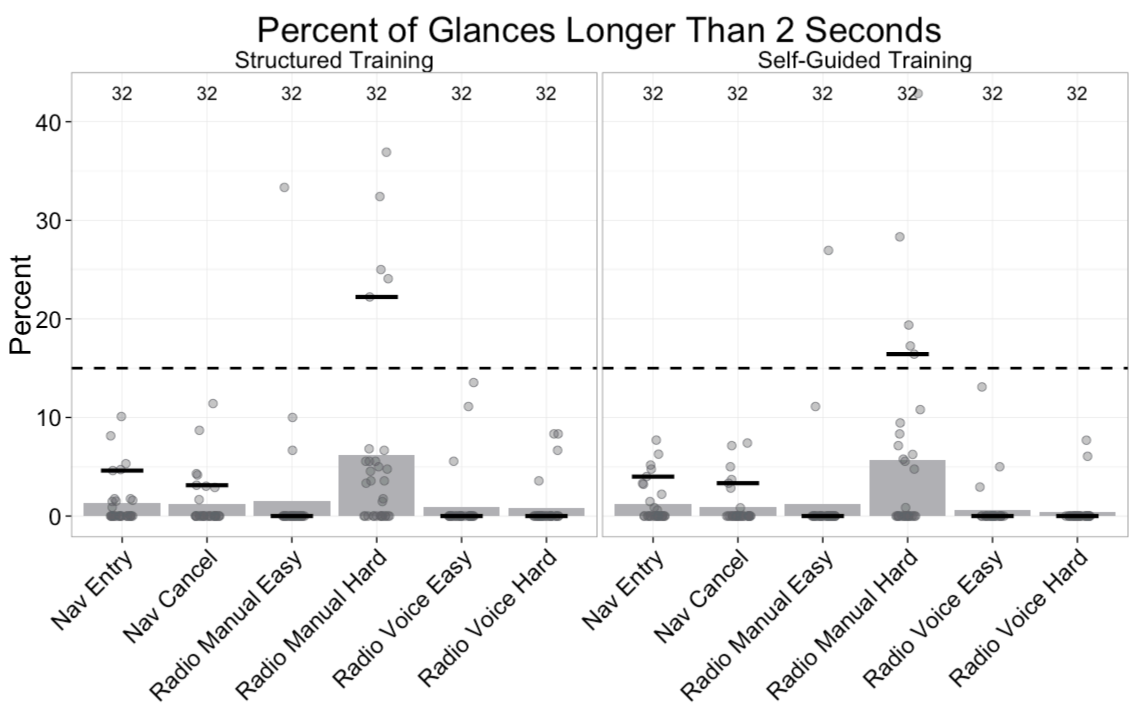


Figure 82: Percentage of single glances longer than 2 seconds task and training condition.

The pattern in the percentage of single glances longer than 2 seconds was not affected by training condition (p = .445). Again note the consistency of the patterning across tasks by training condition.

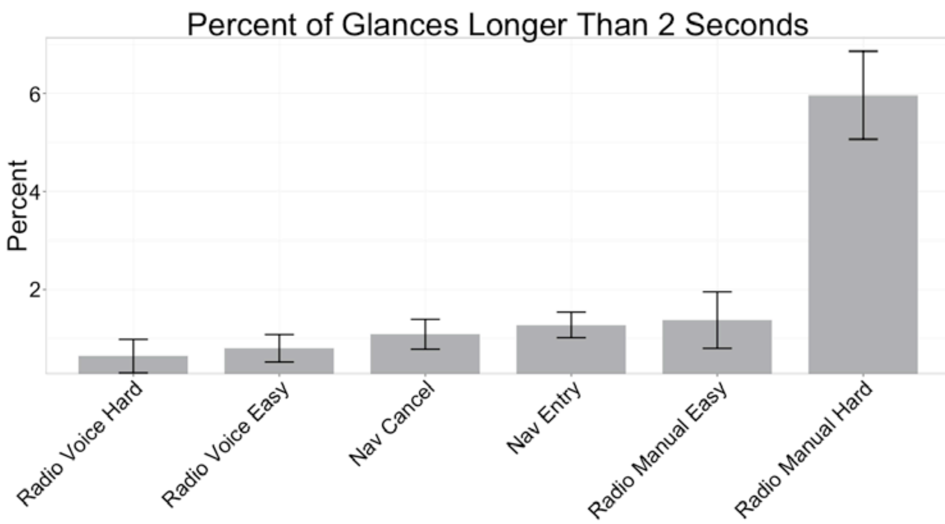


Figure 83: Percentage of single long duration glances collapsed across training conditions showing means and adjusted standard errors.

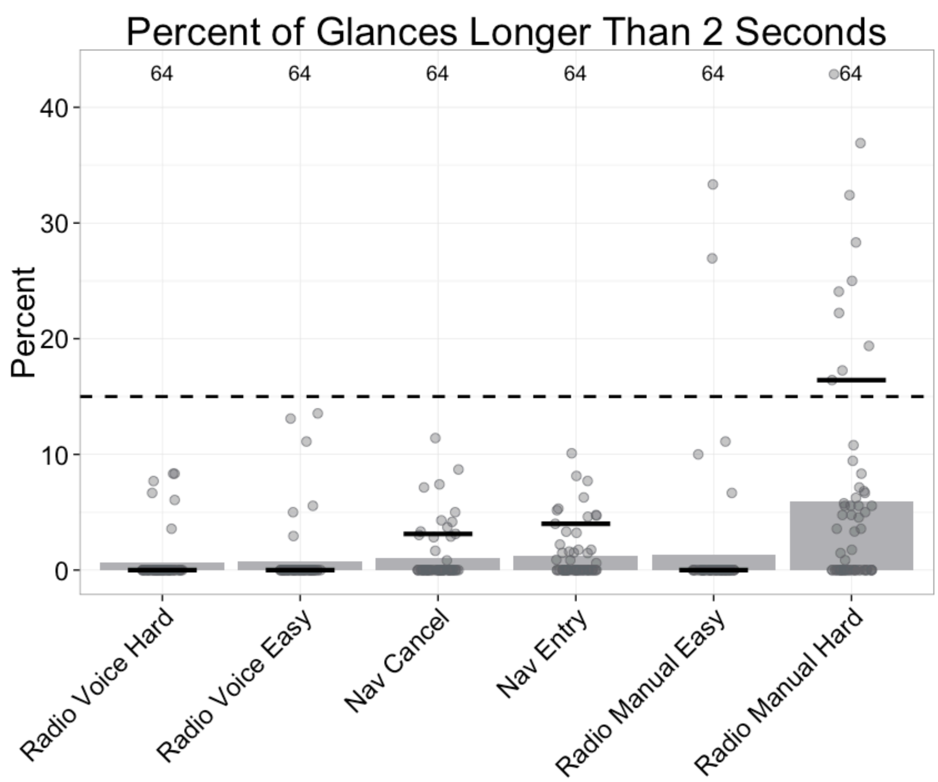


Figure 84: Percentage of single long duration glances with bars representing means and solid circles individual participants.

The percentage of single glances longer than 2 seconds was generally low across the tasks with the exception of the Radio Manual Hard task (visual-manual tuning radio reference task), typically well under 2%. Only the Radio Manual Hard task deviates from this pattern, likely because most participants attempted to complete the task with a small number of long duration glances. Although the Radio Manual Hard task is the only one of the tasks that does not meet a 15% long glance criterion, 84.4% of participants met the criterion in this condition, just shy of the required 85%. This suggests that, given a different sample of participants, this task might have met the recommended safety guideline. Similarly, given how close to the threshold this value is, a modest increase in on-road practice for participants who did not have extensive familiarity with classical manual radio tuning would likely produce results meeting the guideline. It should be noted that, in contrast to Study 1, the current sample does conform to NHTS's specified age distribution and the radio task was adjusted to match the NHTSA specification.

Comparison to Study 1

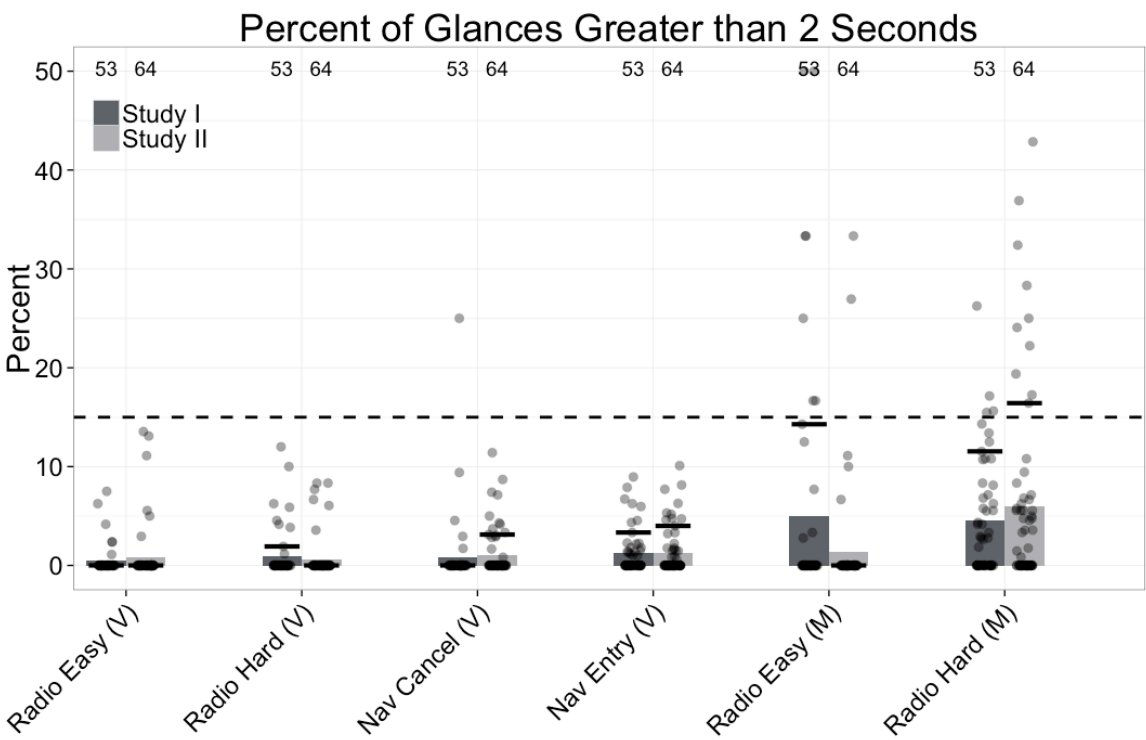


Figure 85: Comparison of percentage of single long duration glances in Study 1 and Study 2.

The percentage of single long duration glance show reasonable similarity across the two studies for both of the voice radio tasks and for voice navigation entry. This consistency indicates that the percentage of single long duration glances metric is not an issue for the voice-command involved interface studies across the tasks considered here.

The modestly higher value seen for the Nav Cancel task in Study 2 may well be related to the previously discussed change in the task in Study 2 that did not provide participants with a direct reminder of the voice command structure to use to complete the task. This resulted in a number of participants finding the task harder to complete and may well have resulted in more relatively long glances to the center console display to look for guidance on how to complete the task. Nonetheless, the 85% bar for the sample is well below the 15% threshold in both Studies, showing high consistency of overall result. The "easy" manual radio tuning task shows the seemingly most overt difference between studies, with values being lower in Study 2. This was likely influenced by Study 1 having a higher proportion of participants in the oldest age category. A greater portion of older participants might reasonably be expected to take longer to visually orient and maintain a single glance duration sufficient to execute the manual control to complete a button press. This suggests that this metric may be particularly sensitive to age considerations in shorter tasks involving this type of interaction. It is less apparent why a higher percentage of long duration glances are observed during the Radio Hard task in Study 2 than in Study 1. One possibility is that the task was changed in the studies (see *Appendix B: Radio Task Changes from Study 1 to 2).* In Study 1, the first step was to press the Vol knob to turn the radio on. In Study 2, the radio was already on as per current NHTSA guidelines and the first step was to press the smaller [RADIO] button (a step not present in Study 1). It may be that the visual orientation and manual button press in Study 2 was slightly more visually demanding with the smaller and less overt control.

Statistical Comparison of Selected Tasks

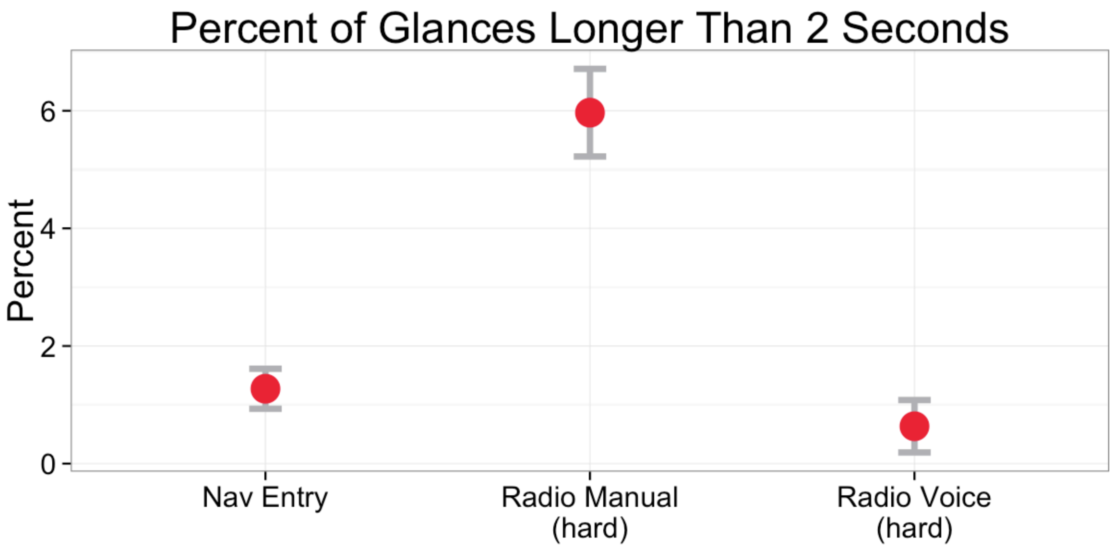


Figure 86: Long glance rate across selected reference tasks.

The percentage of single long duration glances varied significantly across the three reference tasks (X2 = 38.9, $p < .001$, Friedman test). Post-hoc testing reveals that all three tasks are significantly different from one another (Navigation Entry vs. Radio Manual Hard, $p < .001$; Navigation Entry vs. Radio Voice Hard, $p = .023$; Radio Voice Hard vs. Radio Manual Hard $p < .001$). These results parallel the findings for mean glance duration.

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 42: Mean and (SD) of single long glance rate by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	0.94 (2.1)	1.16 (1.7)	2.08 (3.1)	0.91 (2.3)	1.27 (2.3)
Nav Cancel	0.46 (1.3)	1.17 (2.5)	1.36 (3.3)	1.36 (2.2)	1.09 (2.4)
Radio Manual Easy	0.00 (0.0)	0.00 (0.0)	4.46 (10.5)	1.04 (2.9)	1.38 (5.6)
Radio Manual Hard	3.18 (8.1)	4.60 (8.1)	10.89 (14.0)	5.20 (6.9)	5.97 (9.9)
Radio Voice Easy	0.00 (0.0)	0.00 (0.0)	2.51 (4.6)	0.69 (2.8)	0.80 (2.8)
Radio Voice Hard	1.16 (2.6)	0.00 (0.0)	1.00 (2.7)	0.38 (1.5)	0.64 (2.1)
Female	20-24	25-39	40-54	55+	(all)
Nav Entry	0.50 (0.9)	1.18 (1.8)	0.60 (1.7)	0.20 (0.6)	0.62 (1.3)
Nav Cancel	0.54 (1.5)	1.71 (3.3)	0.00 (0.0)	0.35 (1.0)	0.65 (1.9)
Radio Manual Easy	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)	1.25 (3.5)	0.31 (1.8)
Radio Manual Hard	0.91 (2.0)	2.80 (3.2)	0.92 (1.7)	0.45 (1.3)	1.27 (2.3)
Radio Voice Easy	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)
Radio Voice Hard	0.83 (2.4)	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)	0.21 (1.2)
Male	20-24	25-39	40-54	55+	(all)
Nav Entry	1.38 (2.8)	1.13 (1.7)	3.56 (3.6)	1.63 (3.1)	1.93 (2.9)
Nav Cancel	0.38 (1.1)	0.62 (1.2)	2.71 (4.4)	2.37 (2.6)	1.52 (2.8)
Radio Manual Easy	0.00 (0.0)	0.00 (0.0)	8.92 (13.8)	0.83 (2.4)	2.44 (7.7)
Radio Manual Hard	5.44 (11.2)	6.41 (11.1)	20.86 (13.7)	9.96 (6.9)	10.67 (12.2)
Radio Voice Easy	0.00 (0.0)	0.00 (0.0)	5.02 (5.6)	1.39 (3.9)	1.60 (3.9)
Radio Voice Hard	1.49 (3.0)	0.00 (0.0)	2.00 (3.7)	0.76 (2.1)	1.06 (2.6)

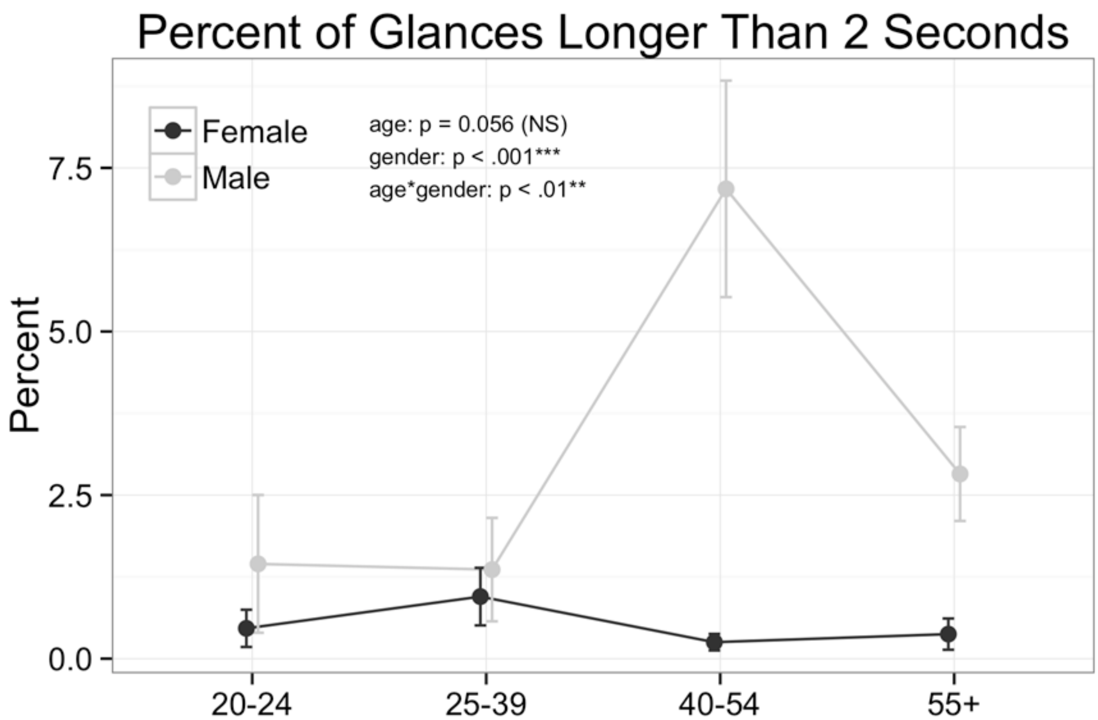


Figure 87: Mean percentage of single glances longer than 2 seconds across age groups and genders.

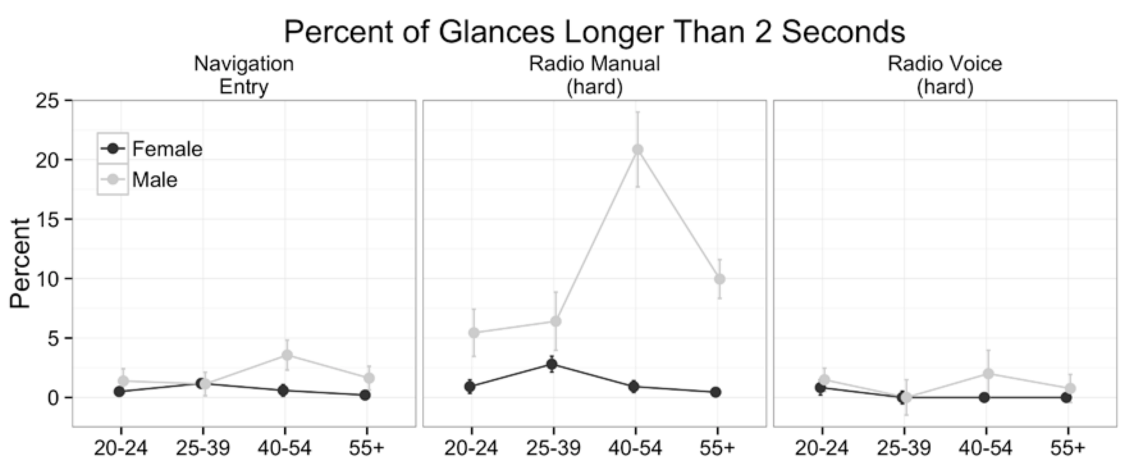


Figure 88: Mean percentage of single glances longer than 2 seconds by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing single long glance rate by age and gender, there were significant effects of gender ($p < .001$) and age * gender interaction ($p = .001$). Long glance duration appears to increase for men over 40. The effect is most pronounced during the Radio Manual Hard task, though it is also nominally evident in both the Navigation Entry and Radio Voice Hard tasks. For women, percentages of long glances remain low regardless of age.

Percentage of Single Long Duration Glances (> 2s) to Device (GTD)

In addition to TEORT glance metrics, Study 1 presented (Reimer et al. 2013: Appendix A) an equivalent series of analyses for GTD metrics. The following section presents an analysis of the percentage of GTD single glances greater than 2 seconds.

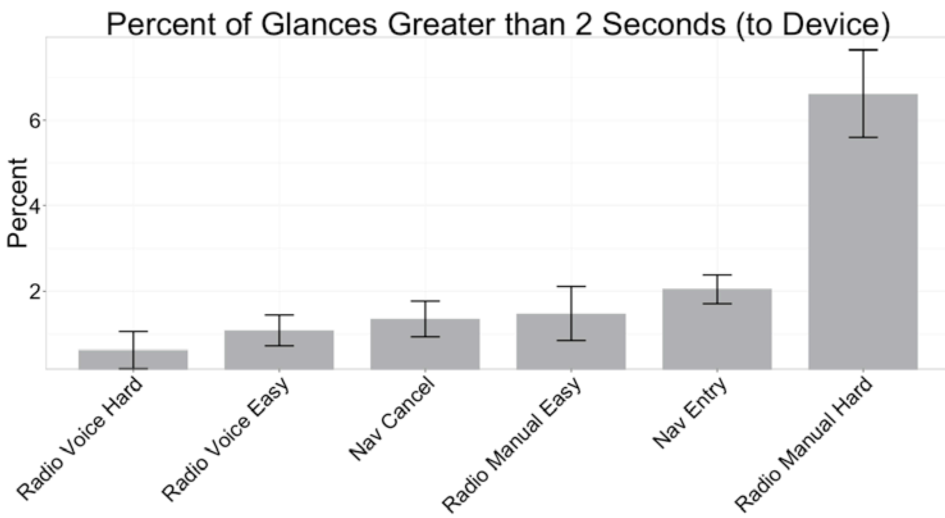


Figure 89: Percentage of single long duration glances (GTD) collapsed across training conditions showing means and adjusted standard errors.

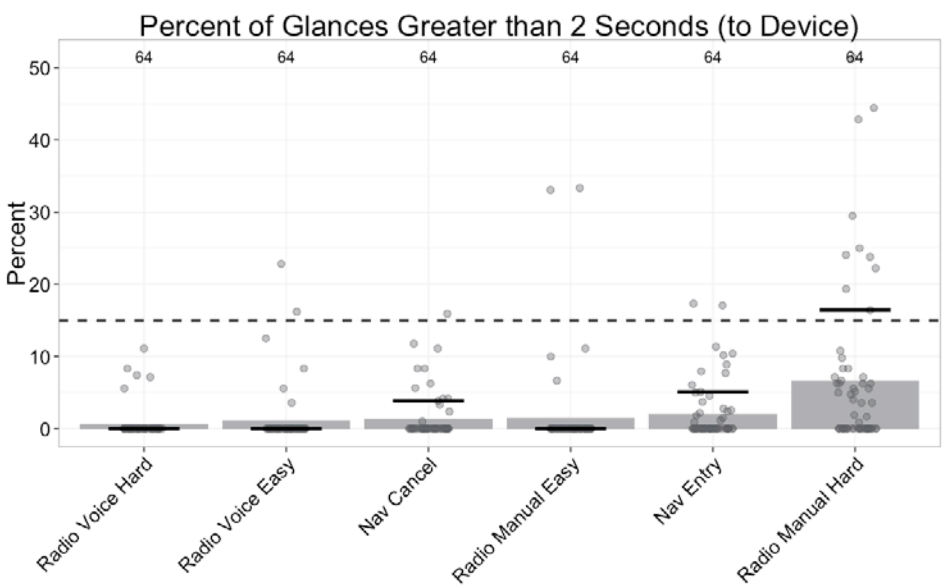


Figure 90: Percentage of single long duration glances (GTD) with bars representing means and solid circles individual participants

The mean percentage of single long duration GTD glances (Figure 89) and the percentage of drivers who showed a high proportion of single long duration GTD glances (Figure 90) was relatively low across the tasks with the exception the Radio Manual Hard task. The long duration glance pattern by task is highly consistent across the GTD and TEORT metrics (see Figure 92). The location of the 85% point for the sample for Radio Manual Hard task is essentially identical using both metrics. This would appear to confirm that the majority of the single long duration TEORT glances observed during this task were to the device itself.

Comparison to Study 1

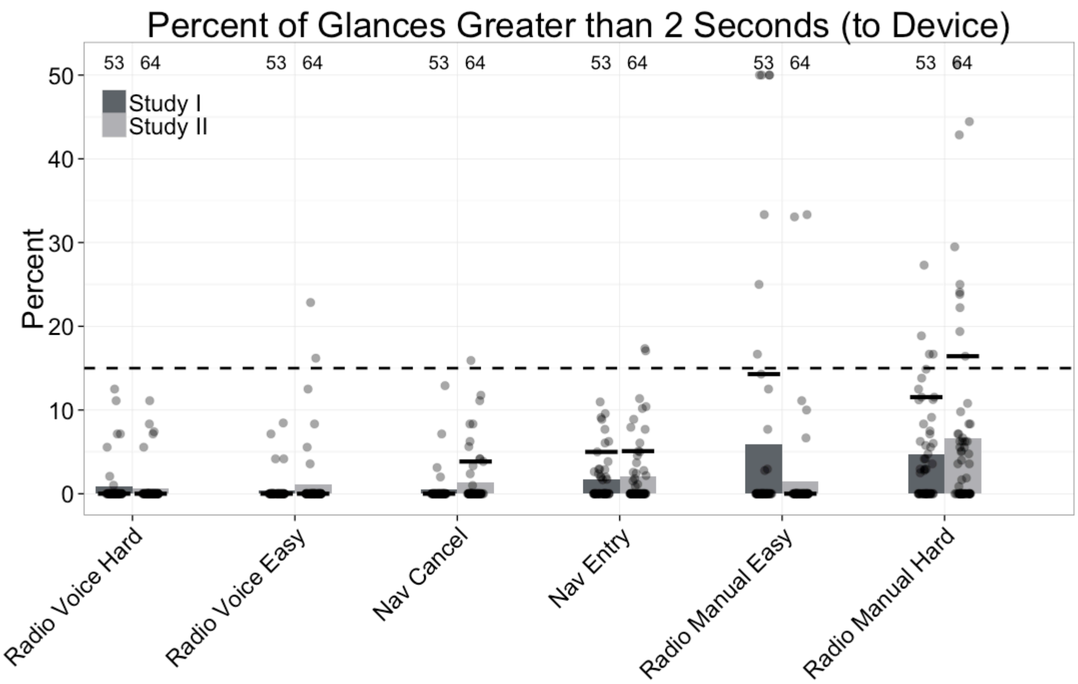


Figure 91: Comparison of percentage of single long glances in Study 1 and Study 2.

GTD glance rate similarity and differences across the two studies are highly similar to what is seen for the TEORT metric presented in the previous section (Figure 85). Please refer to the commentary there.

TEORT vs. GTD Comparison

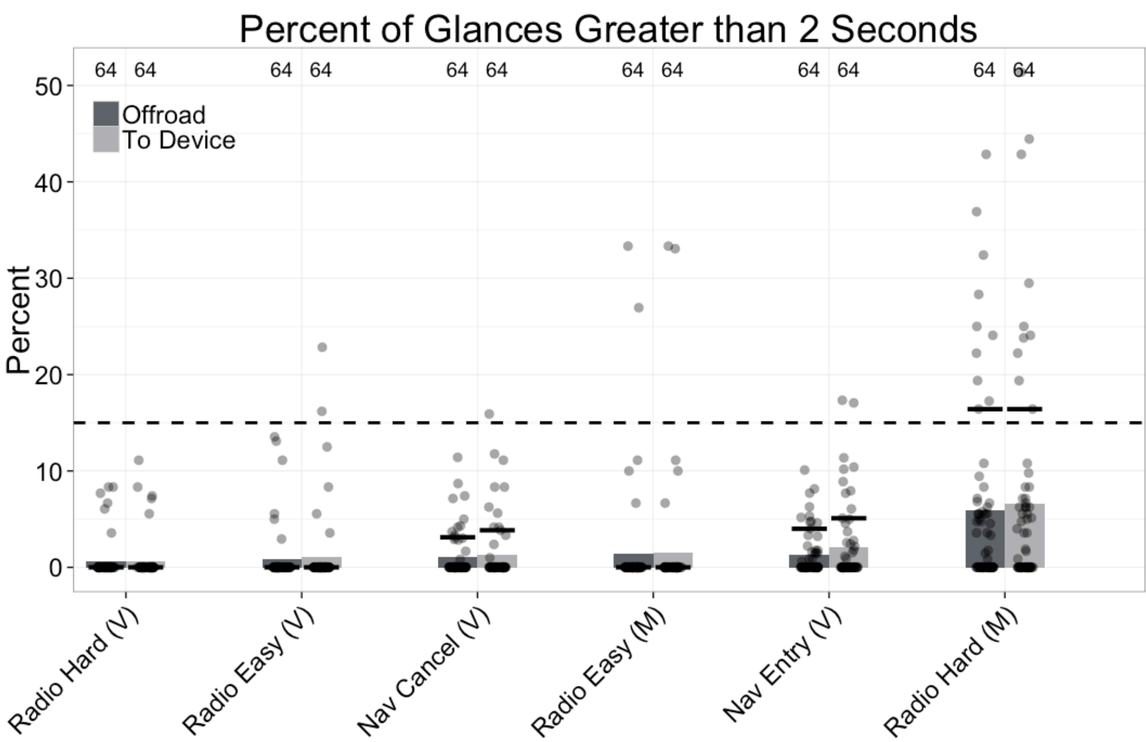


Figure 92: Comparison of percentage of single long glances in Study 1 and Study 2.

As already noted, the TEORT and GT metrics produce fairly consistent results for the percentage of single long duration glances across the tasks.

Statistical Comparison of Selected Tasks

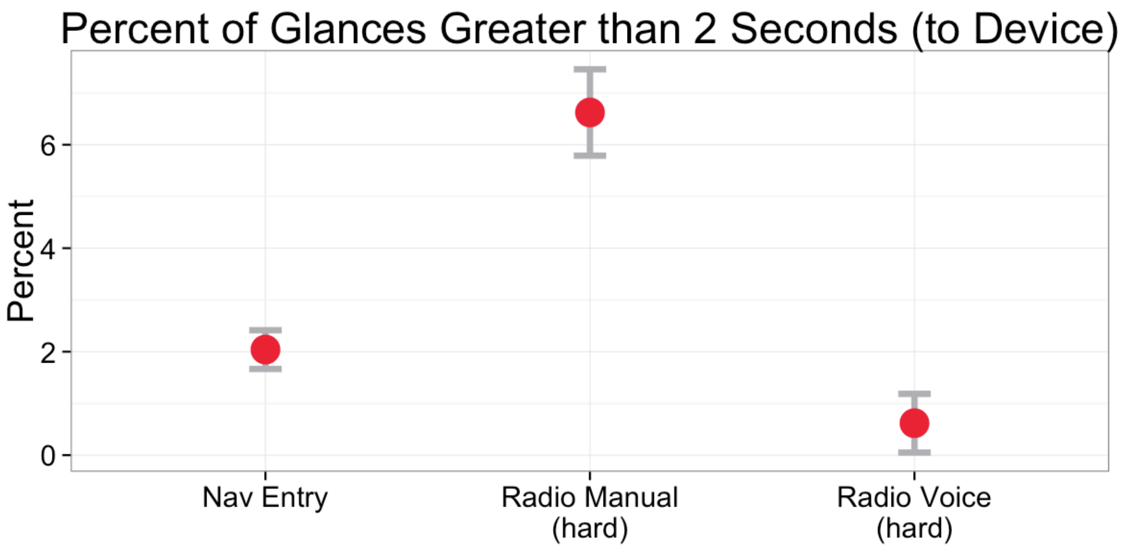


Figure 93: Long glance rate to device across selected reference tasks.

The percentage of long duration glances varied significantly across the three reference tasks (X2 = 34.5, $p < .001$, Friedman test). Post-hoc testing reveals that all three tasks are significantly different from one another (Navigation Entry vs. Radio Manual Hard, $p < .001$; Navigation Entry vs. Radio Voice Hard, $p = .003$; Radio Voice Hard vs. Radio Manual Hard $p < .001$).

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 43: Mean and (SD) of long glance rate by task, broken down by age group and gender.

Combined	20-24	25-39	40-54	55+	(all)
Nav Entry	1.60 (4.3)	1.97 (3.1)	3.27 (5.3)	1.32 (3.2)	2.04 (4.0)
Nav Cancel	0.35 (1.4)	1.52 (3.3)	1.78 (4.4)	1.75 (3.4)	1.35 (3.3)
Radio Manual Easy	0.00 (0.0)	0.00 (0.0)	4.84 (11.4)	1.04 (2.9)	1.47 (6.1)
Radio Manual Hard	4.08 (11.1)	4.73 (8.3)	12.29 (16.3)	5.39 (6.9)	6.62 (11.5)
Radio Voice Easy	0.00 (0.0)	0.00 (0.0)	3.53 (6.8)	0.78 (3.1)	1.08 (3.9)
Radio Voice Hard	1.31 (2.9)	0.00 (0.0)	0.69 (2.8)	0.46 (1.9)	0.62 (2.2)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	0.77 (1.5)	2.09 (3.5)	0.62 (1.8)	0.35 (1.0)	0.96 (2.1)
Nav Cancel	0.70 (2.0)	2.25 (4.4)	0.00 (0.0)	0.00 (0.0)	0.74 (2.5)
Radio Manual Easy	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)	1.25 (3.5)	0.31 (1.8)
Radio Manual Hard	1.13 (2.5)	2.90 (3.3)	1.10 (2.2)	0.50 (1.4)	1.41 (2.5)
Radio Voice Easy	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)
Radio Voice Hard	0.89 (2.5)	0.00 (0.0)	0.00 (0.0)	0.00 (0.0)	0.22 (1.3)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	2.43 (5.9)	1.85 (2.9)	5.91 (6.4)	2.29 (4.3)	3.12 (5.1)
Nav Cancel	0.00 (0.0)	0.78 (1.5)	3.55 (5.8)	3.49 (4.2)	1.96 (3.9)
Radio Manual Easy	0.00 (0.0)	0.00 (0.0)	9.69 (15.0)	0.83 (2.4)	2.63 (8.3)
Radio Manual Hard	7.03 (15.4)	6.55 (11.4)	23.48 (16.6)	10.28 (6.7)	11.84 (14.3)
Radio Voice Easy	0.00 (0.0)	0.00 (0.0)	7.06 (8.4)	1.56 (4.4)	2.16 (5.4)
Radio Voice Hard	1.74 (3.3)	0.00 (0.0)	1.39 (3.9)	0.93 (2.6)	1.01 (2.8)

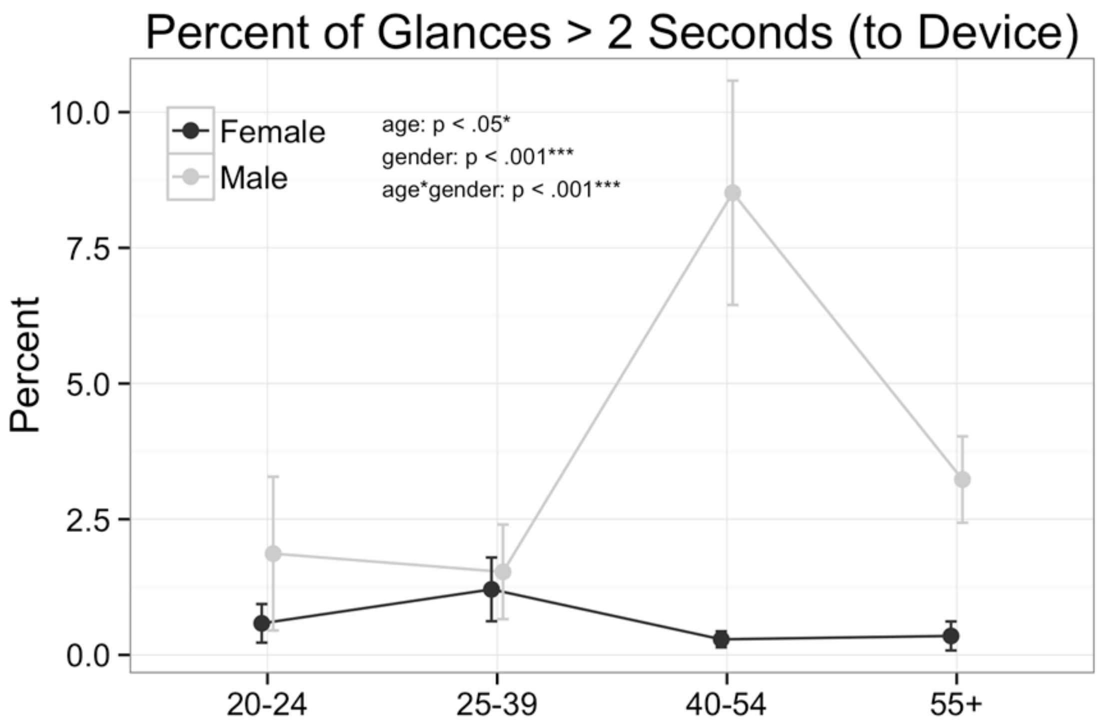


Figure 94: Mean percentage of glances longer than 2 seconds across age groups and genders.

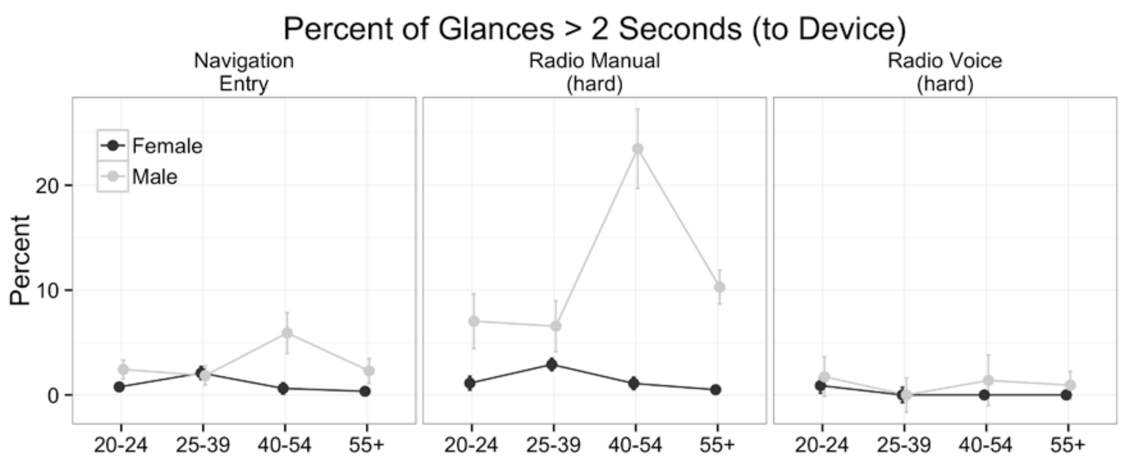


Figure 95: Mean percentage of glances longer than 2 seconds by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing the percentage of single long glances by age and gender, there were significant effects of gender ($p < .001$) and an age * gender interaction ($p = .001$). Long glance duration appears to increase for men over 40. The effect is most pronounced during the Radio Manual Hard task, though it is also nominally evident in the Navigation Entry task. For women, percentages of long glances remain low regardless of age.

Total Eyes-Off-Road Time (TEORT)

Total eyes-off-road time (TEORT) is the most sensitive of the NHTSA metrics to variations in characteristics across the task, and it is correlated with task completion time. The longer a driver takes to complete a task, the more opportunities he/she has to glance off the forward roadway. NHTSA guidelines state that for any given task, a driver's TEORT should not exceed 12s. The extent to which the sample's meet this criterion varied considerably by task.

Table 44: Mean (and standard deviation) of TEORT by training type.

	Structured Training	Self-Guided Training	(combined)
Nav Entry	30.87 (25.5)	29.80 (12.2)	30.33 (19.8)
Nav Cancel	11.43 (8.7)	13.89 (13.3)	12.66 (11.2)
Radio Manual Easy	2.20 (1.6)	2.85 (4.4)	2.52 (3.3)
Radio Manual Hard	13.83 (7.3)	16.16 (11.7)	15.00 (9.7)
Radio Voice Easy	4.98 (4.8)	5.03 (4.1)	5.00 (4.4)
Radio Voice Hard	4.73 (5.6)	4.98 (3.9)	4.86 (4.8)

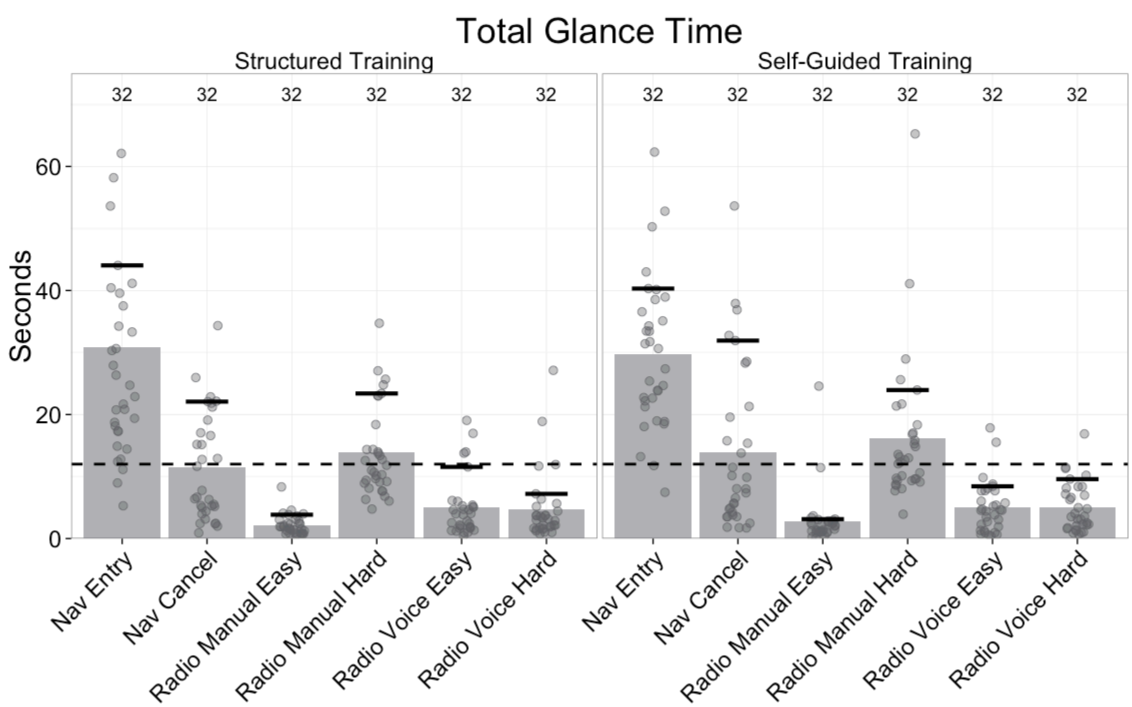


Figure 96: TEORT by task and training condition.

TEORT was not affected by training condition (p = .466). As with task completion time and most other metrics, TEORT was extremely consistent across training conditions and is presented below collapsed across training conditions.

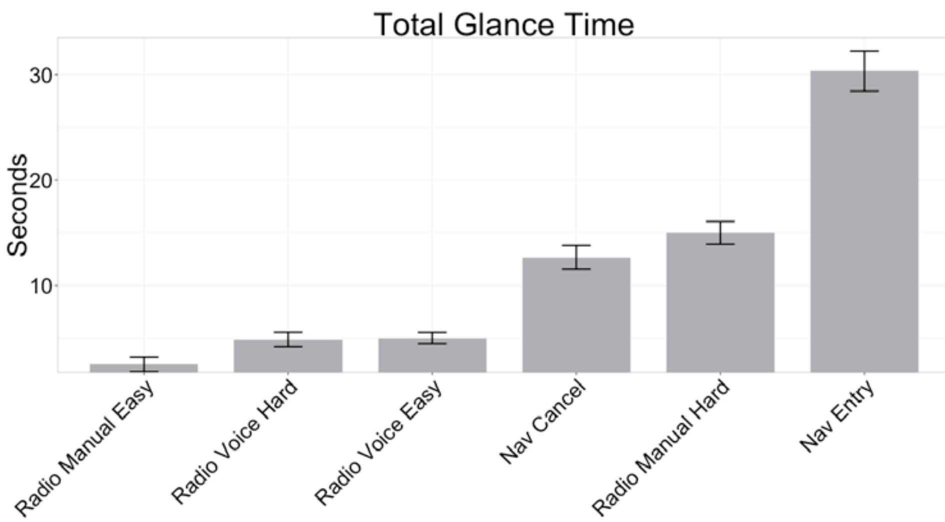


Figure 97: TEORT collapsed across training conditions showing means and adjusted standard errors.

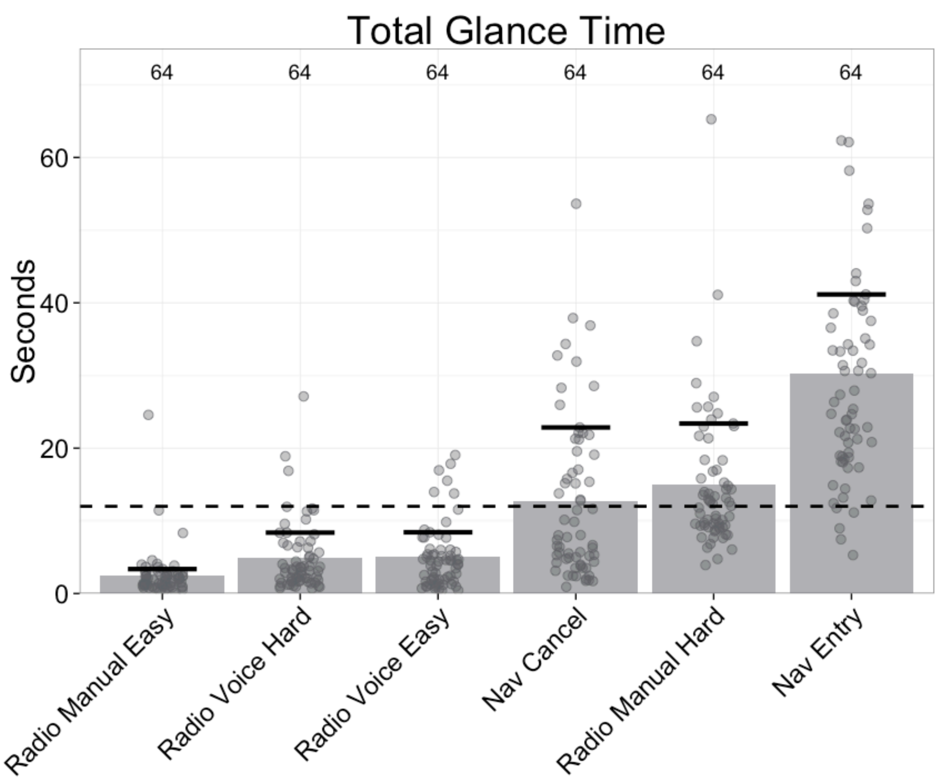


Figure 98: TEORT with bars representing means and solid circles individual participants.

Completion of the Navigation Entry task was associated with a large cumulative off-road glance time. Only 7.8% of the sample was able to complete the task with less than 12s of TEORT, well below the 85% threshold specified by NHTSA's guidelines for visual-manual DVI tasks. This finding is highly consistent with the results of Study 1. Also consistent with Study 1 was the finding that the Radio Manual Hard task did not meet criterion for the 12 second threshold. It should again be noted that the current sample does conform to NHTS's specified age distribution and the radio task was adjusted to match the NHTSA specification for the visual-manual radio tuning reference task.

Comparison to Study 1

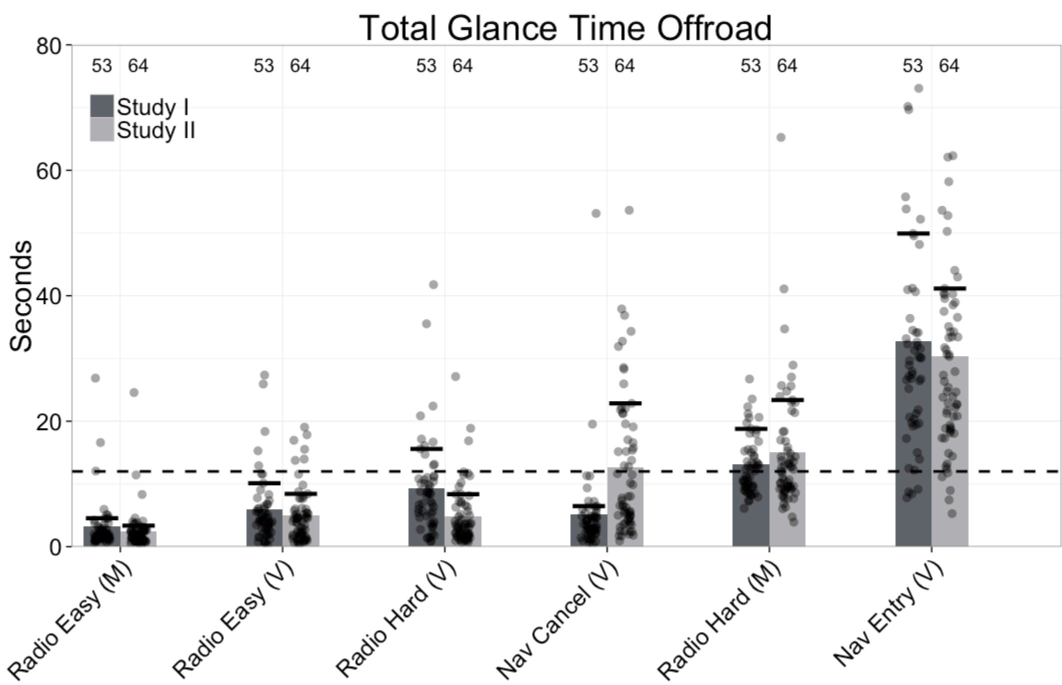


Figure 99: Comparison of TEORT during tasks common to Study 1 (dark bars) and Study 2 (light bars).

The Nav Cancel task also failed to attain an 85% pass rate; the observed percentage was 59.4%. This contrasts with the findings from Study 1 where this task was completed with a relatively brief TEORT. The longer glance time is likely attributable to the more difficult form of the task for this sample (i.e. no direct prompting on how to complete the task) which may have resulted in more reliance on looking to the display for support on how to complete the task.

It can be observed that the multi-step manual radio tuning task (Radio Hard) also shows a somewhat higher TEORT value in Study 2. In Study 1, this task required pressing the large Vol knob to turn of the radio, touching a touch-screen soft key to change the band, and manual rotation of the tuning knob to find a specified station. As detailed in the *Methods* and in *Appendix B*, the manual Radio Hard task was modified in Study 2 to explicitly conform to NHTSA's (2012) specification of a manual radio turning reference task. The revised task started with the radio already on and the steps involved to complete the task consisted on pressing a mode button, [RADIO], to make the band selections visible, then touching a touch-screen soft key to change the band, and manual rotation of the tuning knob to find a specified station. The [RADIO] mode button is somewhat smaller, has a lower profile, and is located to the right of the Vol control (Figure 5). This size and positioning may have contributed to it taking somewhat longer for drivers to locate and manually engage.

Except for the Nav Cancel task and the Manual Radio Hard task, it can be observed that the remaining tasks showed either nominally lower or markedly lower mean TEORT means in Study 2. One possible explanation may lie in Study 2 having a lower percentage of participants in the 60-69 age range. TEORT breakdowns by age in Study 1 found that the older age group showed larger values on this metric and the age breakdown for Study 2 shows a similar finding (Figure 101).

An additional consideration for the apparently lower TEORT for the Radio Verbal Hard condition in this second sample may be found in the change in the initial DVI state in which the task was initiated. Speciffically, the task was initiated when the radio was already on to conform to NHTSA's (2012) specification. This reduced by a minimum of one the number of commands the driver need to issue to complete the task. Reducing the number of voice-command steps appeared to reduce the number of glances (Figure 118) to the display, which some drivers were observed to engage in even when they did not have to view the display to acquire specific information. Reducing the number of glances corresponds here to lower TEORT.

Statistical Comparison of Selected Tasks

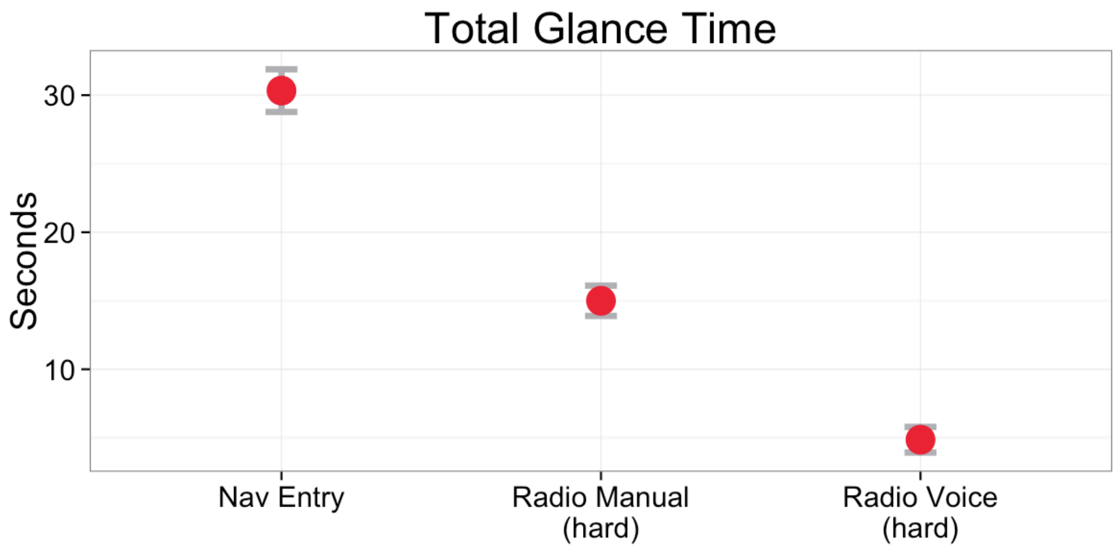


Figure 100: TEORT across selected reference tasks.

TEORT differed significantly across the three reference tasks ($X^2 = 105.5$, $p < .001$, Friedman test). Post-hoc testing revealed that all three tasks were significantly different from one another ($p < .001$ for all comparisons, Wilcoxon signed rank tests). Interestingly, while the Radio Manual Hard task failed to meet the NHTSA criterion, the equivalent Radio Voice Hard task was within the criterion (95.3% of the sample completed the task in less than 12s of TEORT).

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 45: Mean and (SD) of TEORT by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	27.02 (12.0)	25.79 (11.7)	30.91 (15.9)	37.61 (32.1)	30.33 (19.8)
Nav Cancel	13.71 (15.2)	9.06 (8.2)	11.51 (8.3)	16.35 (11.3)	12.66 (11.2)
Radio Manual Easy	1.90 (1.9)	1.73 (0.9)	3.22 (5.8)	3.25 (2.4)	2.52 (3.3)
Radio Manual Hard	10.96 (6.8)	15.04 (14.2)	14.70 (5.5)	19.29 (9.2)	15.00 (9.7)
Radio Voice Easy	2.74 (2.0)	4.72 (3.4)	5.74 (5.6)	6.80 (5.2)	5.00 (4.4)
Radio Voice Hard	5.42 (6.7)	4.62 (4.6)	4.48 (4.7)	4.91 (2.9)	4.86 (4.8)
Female	20-24	25-39	40-54	55+	(all)
Nav Entry	25.97 (14.7)	26.42 (11.7)	20.84 (10.5)	25.20 (10.6)	24.60 (11.6)
Nav Cancel	19.07 (19.5)	9.63 (10.0)	13.41 (8.8)	12.99 (10.5)	13.78 (12.8)
Radio Manual Easy	2.33 (2.5)	1.90 (0.9)	1.42 (0.4)	3.66 (3.4)	2.33 (2.2)
Radio Manual Hard	12.57 (9.2)	13.73 (5.6)	15.81 (6.1)	16.89 (8.8)	14.75 (7.4)
Radio Voice Easy	3.26 (2.5)	6.37 (3.4)	2.55 (1.5)	4.82 (2.8)	4.25 (2.9)
Radio Voice Hard	4.03 (3.3)	3.35 (1.9)	2.55 (1.5)	4.66 (2.6)	3.65 (2.4)
Male	20-24	25-39	40-54	55+	(all)
Nav Entry	28.08 (9.5)	25.16 (12.3)	40.98 (14.1)	50.02 (41.7)	36.06 (24.4)
Nav Cancel	8.35 (7.0)	8.50 (6.6)	9.61 (8.0)	19.72 (11.7)	11.55 (9.5)
Radio Manual Easy	1.47 (1.1)	1.55 (0.9)	5.02 (8.0)	2.83 (1.1)	2.72 (4.2)
Radio Manual Hard	9.35 (2.7)	16.36 (19.9)	13.59 (4.9)	21.69 (9.6)	15.24 (11.7)
Radio Voice Easy	2.23 (1.4)	3.07 (2.6)	8.94 (6.4)	8.78 (6.4)	5.75 (5.5)
Radio Voice Hard	6.81 (8.9)	5.89 (6.2)	6.42 (6.0)	5.16 (3.3)	6.07 (6.2)

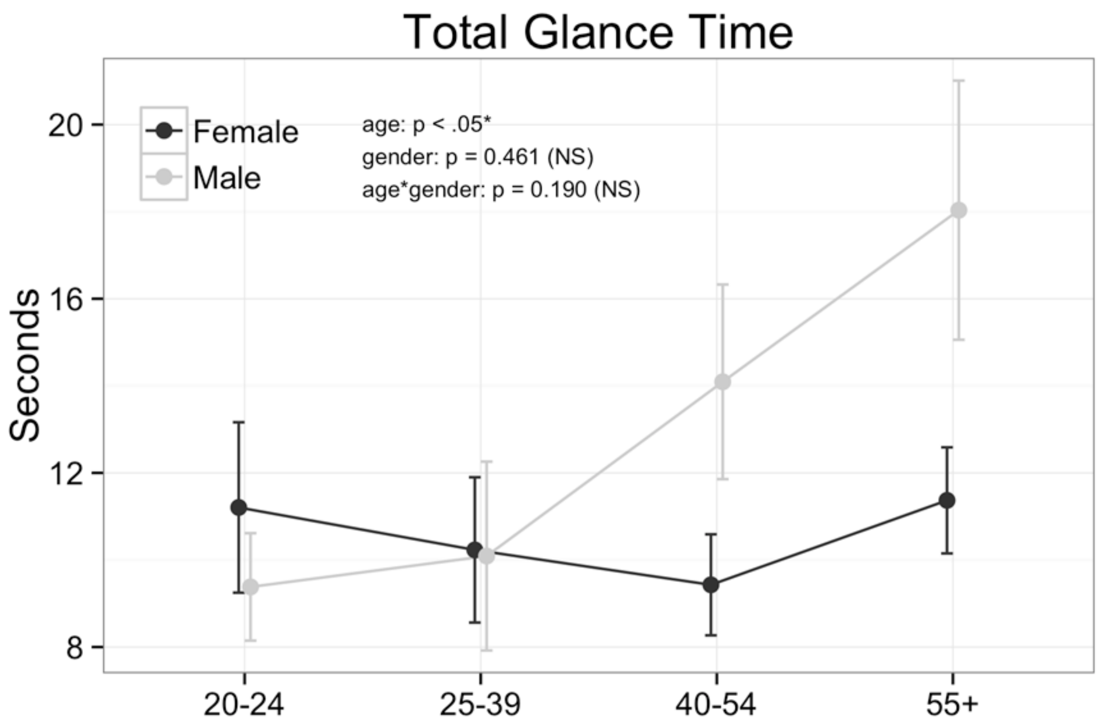


Figure 101: TEORT across age groups and genders.

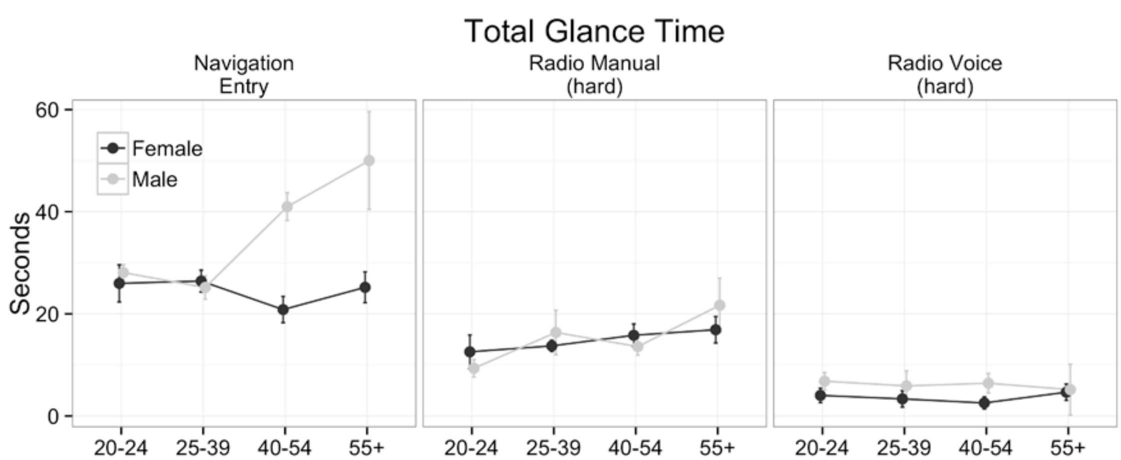


Figure 102: TEORT by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing TEORT by age and gender, there is a significant effect of age ($p = .046$). This effect is driven almost exclusively by men older than 40 completing the Navigation Entry task.

Total Glance Time to Device (The Alliance Metric)

While NHTSA's recent guidelines (2013) assess glance behavior in terms of the total time a driver's eyes are directed away from the forward roadway (TEORT metric), the earlier Alliance (2006) guidelines consider the total time during a task that a driver's eyes are directed to a device, and specifies a 20 second criteria. This section considers glance time in terms of total glance time to the device (GTD).

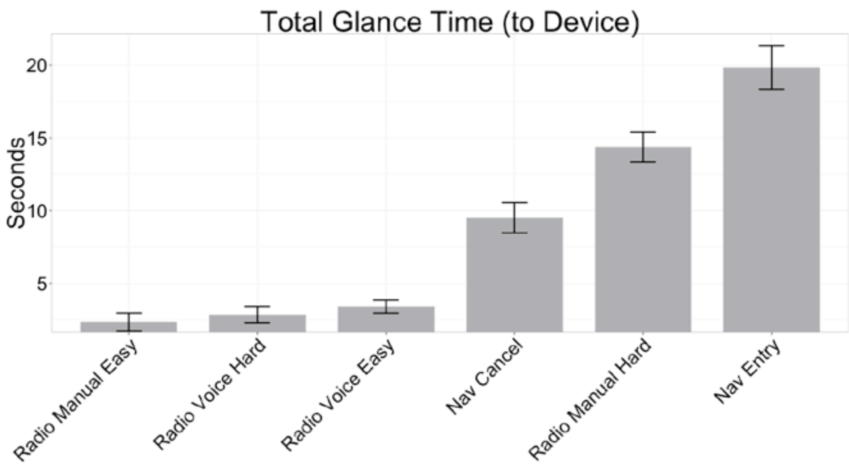


Figure 103: Means and standard deviations of total glance time to device (GTD).

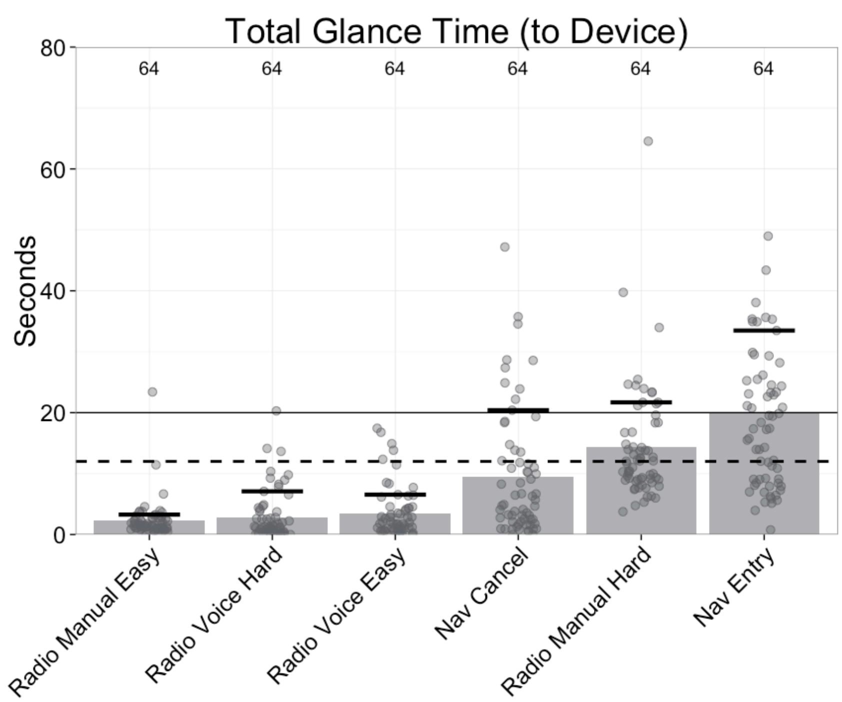


Figure 104: GTD collapsed across training conditions. The thin horizontal line at 20 seconds represents the Alliance of Automobile Manufacturers' criterion 2.1A for total glance time. The dashed line again shows the 12 second threshold for TEORT set by NHTSA.

As in the TEORT analysis in the previous section, the Navigation Entry task required the greatest total glance time, followed by the Radio Manual Hard task. The values presented here represent the mean of two trials of each task employing a sample following NHTSA's guidelines for age distribution and balancing by gender. If the Alliance of Automobile Manufacturer's' criterion that 85% of the sample should complete a task within 20 seconds were applied, the Nav Cancel, Radio Manual Hard tuning task, and the full destination address entry task all fall outside this threshold. Given that Nav Cancel and Radio Manual Hard tuning fall fairly close to the threshold, it is likely that additional experience with these tasks might well bring performance within the guidelines. In the case of the full address entry task, however, this seems less likely as the 85% point for the sample falls in excess of 50% above the threshold (see Table 47).

Comparison to Study 1

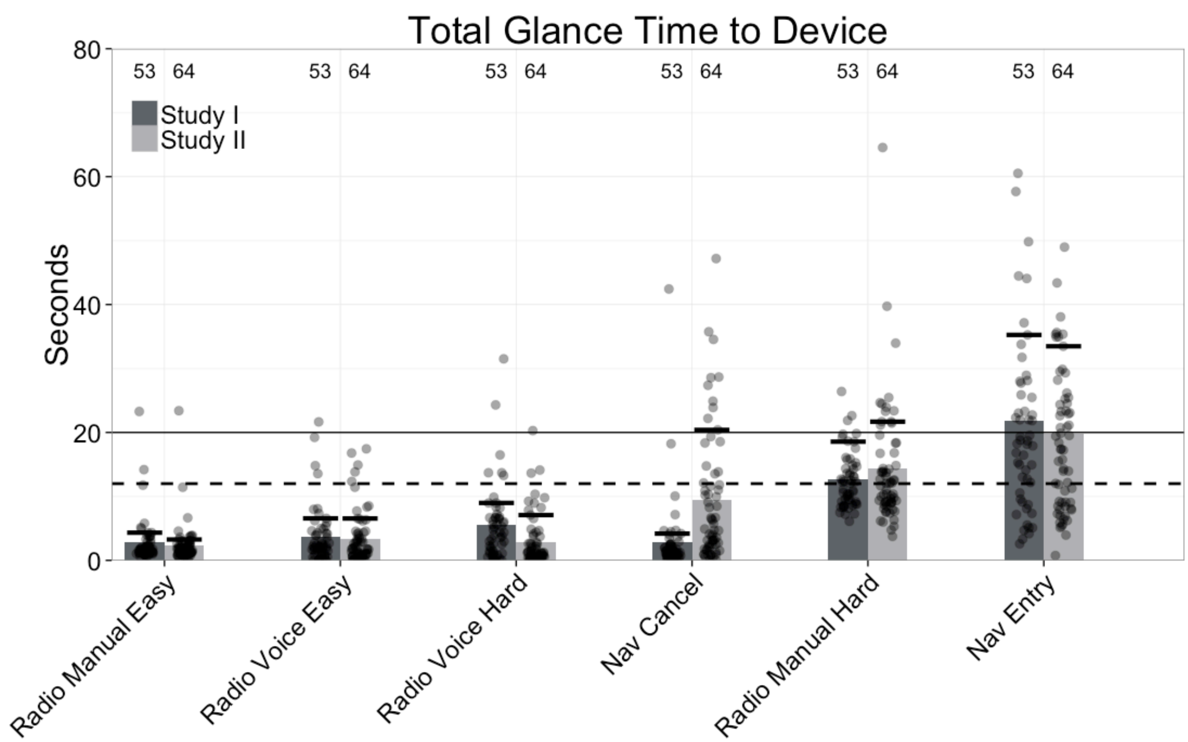


Figure 105: Comparison of total glance time to device in Study 1 and Study 2.

Glance rates were similar across the two studies for the two manual radio tuning tasks and the voice based radio pre-set request (radio easy) task. This is understandable as the tasks were the same in both studies. As was the case for TEORT, total GTD time increased notably in Study 2 for the Nav Cancel task, likely because this version of the task was not preceded by explicit instructions on how to perform it. We hypothesize that Study

1 actually underestimated the visual demands of the Nav Cancel task for new users because of the explicit instruction provided on how to complete the task.

Please see the discussion around the *Comparison to Study 1* section of *Total Eyes-Off-Road Time (TEORT*) (Figure 99) for further consideration of the modest differences observed between Study 1 and 2 on total glance time. It can be noted that while the patterning is similar for both metrics (TEORT and GTD), the apparent magnitude of the differences between studies appears to be somewhat greater for the TEORT metric than for GTD. In essence, assessments based upon TEORT may be more variable than assessments based upon GTD. There may be some value in further consideration of the relative consistency of each metric across assessment samples to more effectively understand the utilization of each measure.

TEORT vs. Glace to Device Comparison

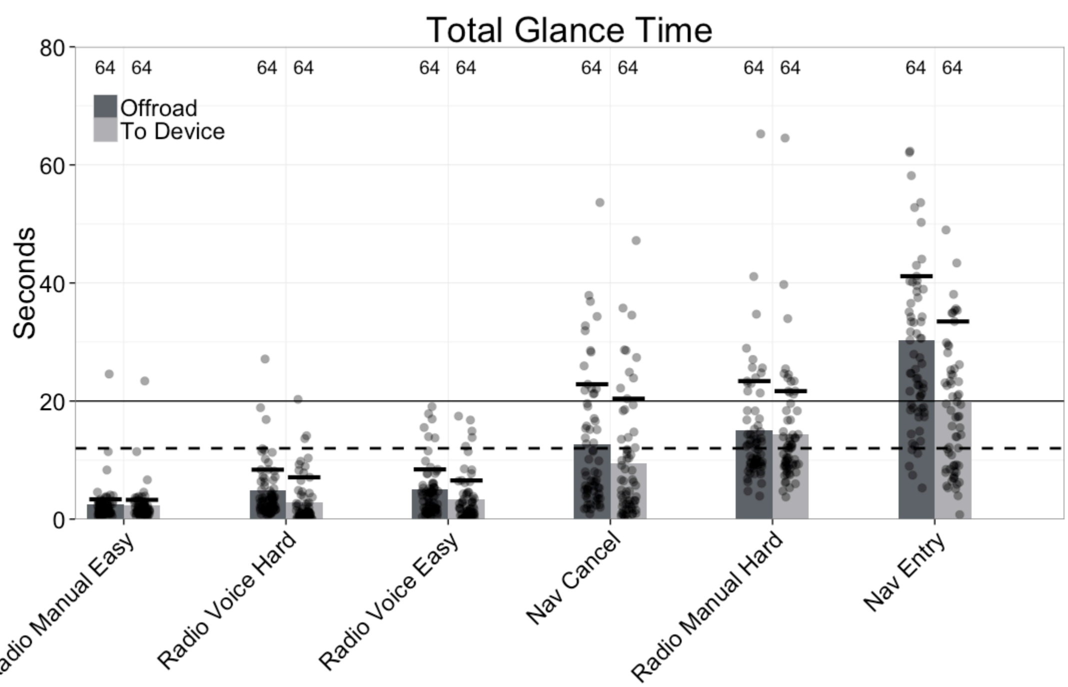


Figure 106: Comparison of total glance time to device and off-road in general (TEORT).

As might be expected, it can be observed in the figure above that for most tasks the GTD metric produces nominally lower values than the TEORT metric. This is most apparent in the full address entry navigation task. Given the relatively long overall duration of the Nav Entry task (M 113s; SD 13s), drivers may be more prone to engage in other driving related glances such as checking speed and inspection of rear and side mirrors.

Statistical Comparison of Selected Tasks

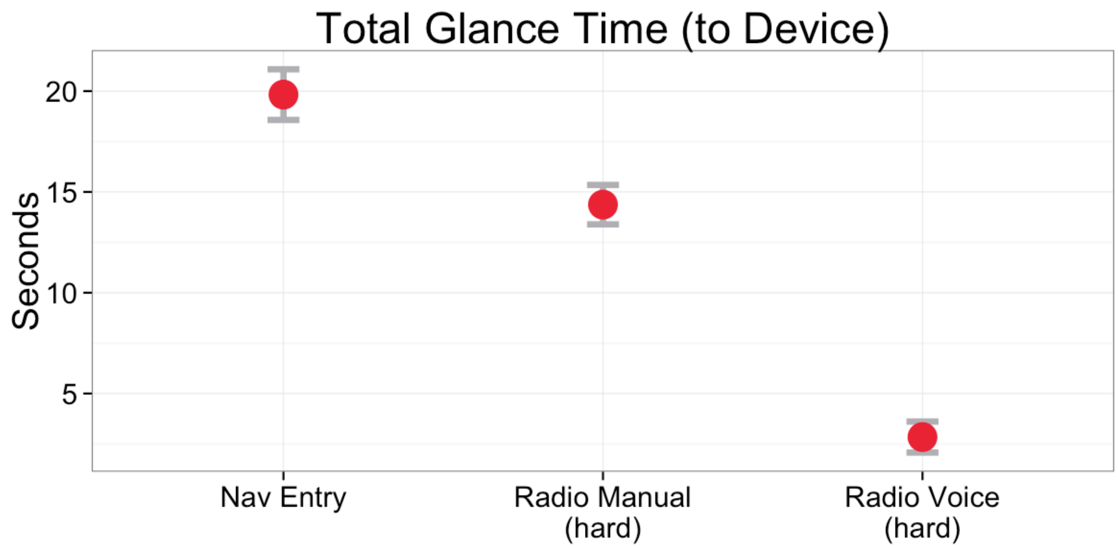


Figure 107: Total glance time to device across selected reference tasks.

Total GTD time varied significantly across the three reference tasks ($X^2 = 93.2$, $p < .001$, Friedman test) with total GTD being marked lower for the voice-command method for station specific radio tuning (radio "hard" task). Post-hoc testing reveals that all three tasks are significantly different from one another (Navigation Entry vs. Radio Manual Hard, p = .008; Navigation Entry vs. Radio Voice Hard, p < .001; Radio Voice Hard vs. Radio Manual Hard p < .001).

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 46: Mean and (SD) of total GTD time by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	17.27 (9.1)	17.87 (9.5)	19.74 (12.3)	24.41 (26.2)	19.83 (15.7)
Nav Cancel	11.04 (14.2)	7.06 (7.1)	8.60 (7.7)	11.34 (11.0)	9.51 (10.3)
Radio Manual Easy	1.66 (1.5)	1.61 (0.9)	3.01 (5.5)	3.02 (2.5)	2.32 (3.2)
Radio Manual Hard	10.49 (6.6)	14.80 (14.1)	14.21 (5.1)	17.97 (8.8)	14.37 (9.4)
Radio Voice Easy	1.47 (1.6)	3.24 (3.1)	4.37 (5.3)	4.52 (5.4)	3.40 (4.2)
Radio Voice Hard	3.47 (5.5)	2.62 (3.8)	2.71 (4.1)	2.56 (2.8)	2.84 (4.1)
Female	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	16.79 (11.8)	17.74 (8.7)	11.45 (9.1)	13.87 (11.1)	14.96 (10.0)
Nav Cancel	16.03 (18.2)	7.58 (8.7)	9.81 (8.7)	7.24 (9.3)	10.16 (11.9)
Radio Manual Easy	2.03 (2.0)	1.73 (0.9)	1.20 (0.4)	3.39 (3.5)	2.09 (2.1)
Radio Manual Hard	12.30 (9.0)	13.39 (5.7)	15.24 (5.8)	15.28 (7.8)	14.05 (7.0)
Radio Voice Easy	1.90 (2.0)	4.53 (3.3)	1.30 (1.5)	2.43 (3.0)	2.54 (2.7)
Radio Voice Hard	2.56 (3.2)	1.37 (1.3)	0.79 (1.0)	1.75 (2.5)	1.62 (2.2)
Male	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	17.75 (6.1)	18.01 (10.8)	28.04 (9.3)	34.96 (33.0)	24.69 (18.8)
Nav Cancel	6.05 (6.5)	6.54 (5.6)	7.39 (6.9)	15.45 (11.5)	8.86 (8.5)
Radio Manual Easy	1.29 (1.0)	1.50 (0.9)	4.82 (7.6)	2.65 (1.1)	2.56 (4.0)
Radio Manual Hard	8.68 (2.4)	16.22 (19.7)	13.18 (4.5)	20.67 (9.3)	14.69 (11.5)
Radio Voice Easy	1.05 (1.2)	1.96 (2.3)	7.44 (6.0)	6.60 (6.6)	4.26 (5.3)
Radio Voice Hard	4.37 (7.2)	3.88 (5.1)	4.63 (5.2)	3.37 (3.1)	4.06 (5.1)

Table 47: Percentage of participants who would pass The Alliance total glance time to device criteria, if applied to the tasks assessed. Tasks that do not meet the 85% threshold under the assessment protocol employed in this study are highlighted in red.

	Total Glance Time to Device
Nav Entry	**57.80%**
Nav Cancel	**84.40%**
Radio Manual Easy	98.40%
Radio Manual Hard	**79.70%**
Radio Voice Easy	100.00%
Radio Voice Hard	98.40%

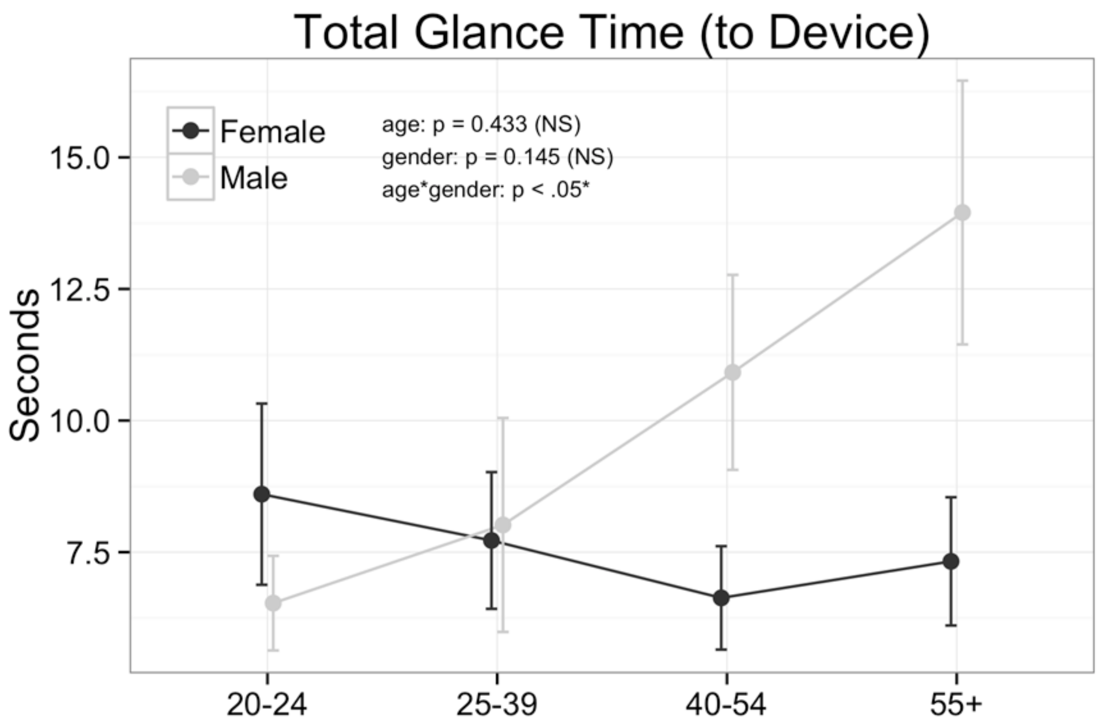


Figure 108: Mean total GTD time across age groups and genders.

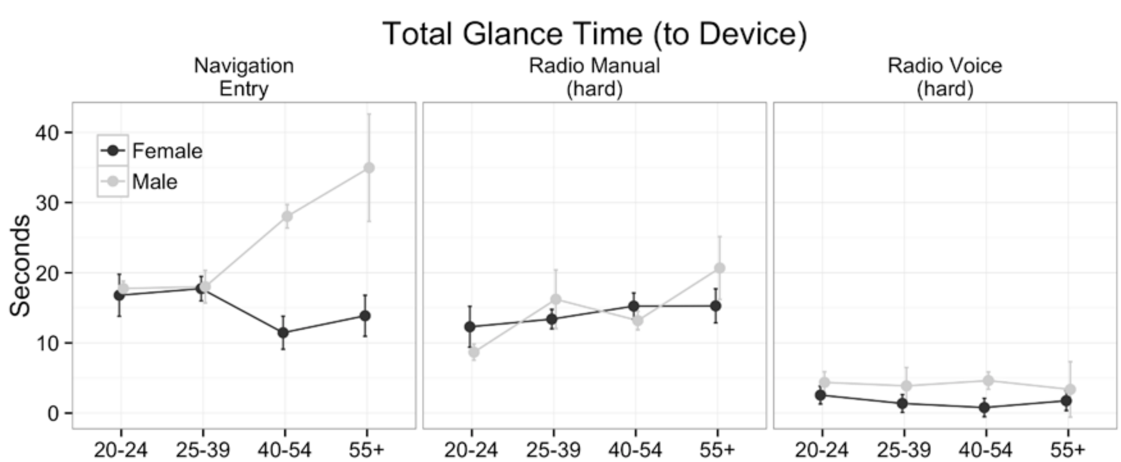


Figure 109: Mean total GTD time by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Decomposing total GTD time by age and gender, there was a significant age * gender interaction (p = .040). GTD time appears to increase for men over 40. The effect is most evidently driven by the Nav Entry task. For women, GTD time is largely consistent as a function of age.

Number of Off-Road Glances

Although the number (raw frequency) of off-road glances is not part of NHTSA's guidelines for visual-manual distraction, this metric is directly relevant to a consideration of TEORT. Therefore, we present raw glance frequency in the data below.

Table 48: Mean (and standard deviation) of number of glances by training type.

	Structured Training	Self-Guided Training	(combined)
Nav Entry	35.25 (23.2)	36.58 (14.6)	35.91 (19.3)
Nav Cancel	13.05 (8.2)	16.24 (15.1)	14.65 (12.1)
Radio Manual Easy	2.53 (1.8)	3.25 (3.8)	2.89 (3.0)
Radio Manual Hard	13.16 (7.0)	16.07 (8.0)	14.61 (7.6)
Radio Voice Easy	5.90 (4.3)	6.73 (4.7)	6.32 (4.5)
Radio Voice Hard	5.95 (5.9)	6.81 (4.9)	6.38 (5.4)

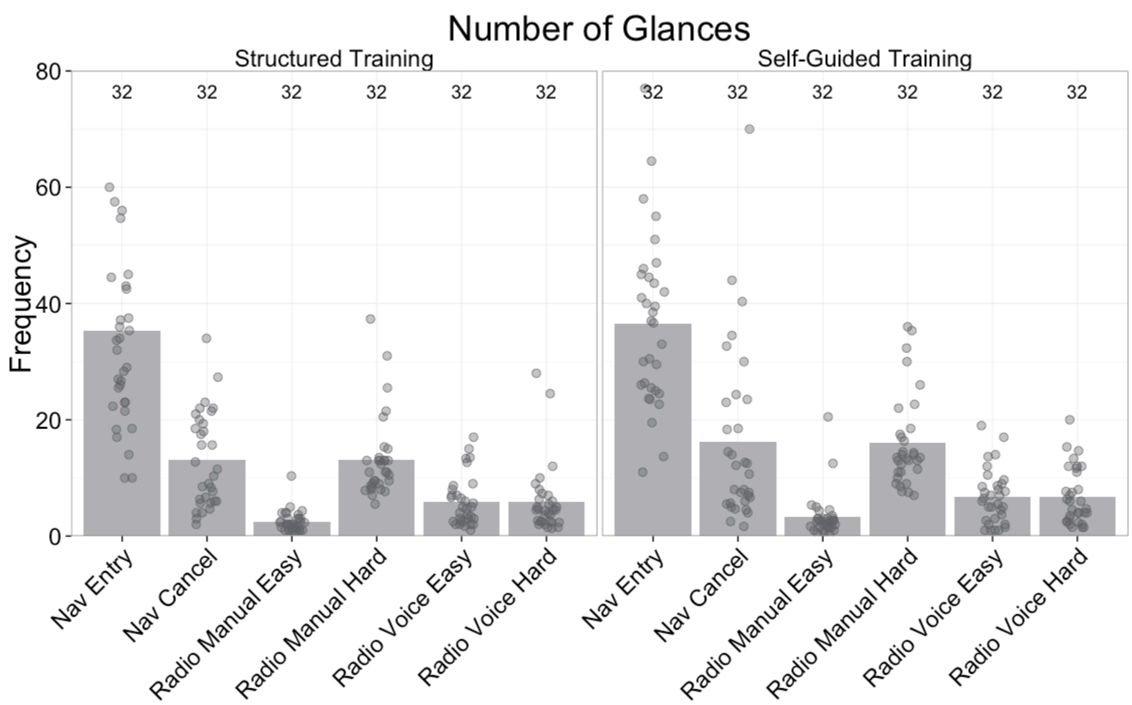


Figure 110: Mean number of glances by task and training condition.

The number of glances per task was not affected by training condition ($p = .219$). As for the other glance metrics, note the consistency of patterning by task across training conditions.

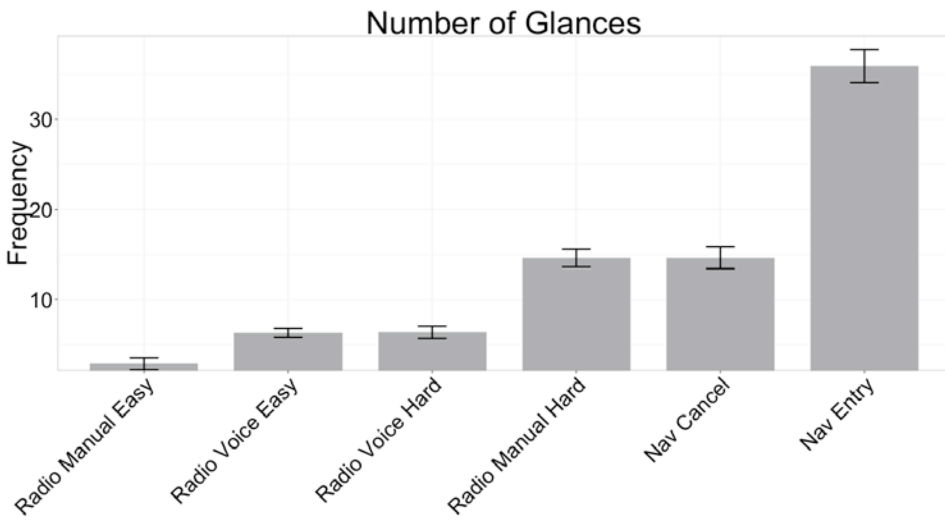


Figure 111: Number of glances showing means and adjusted standard errors.

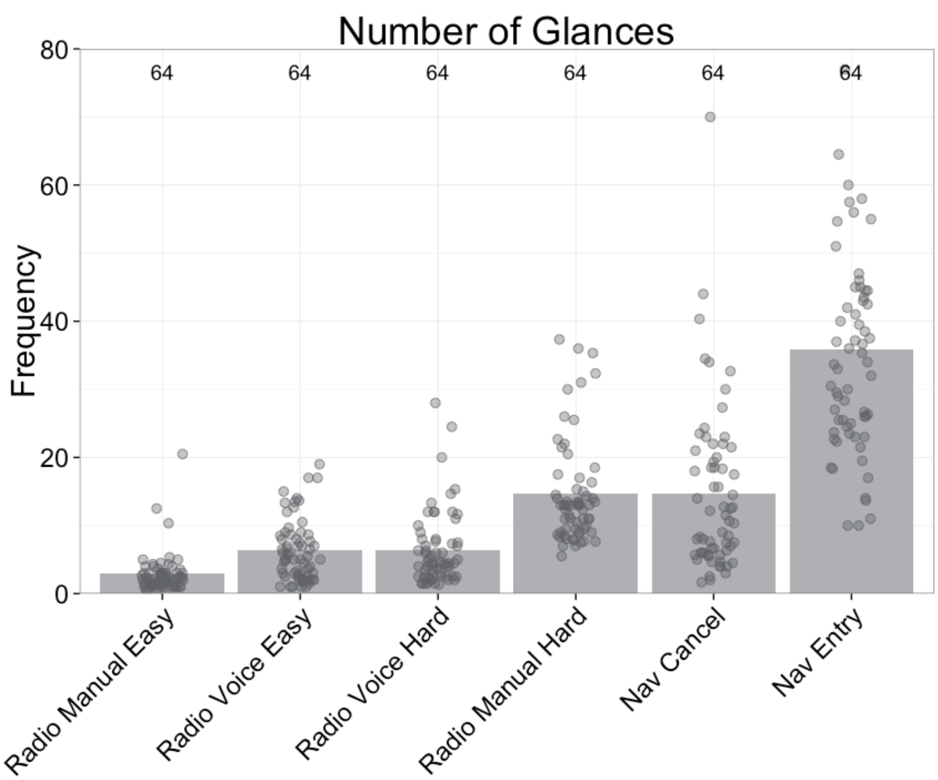


Figure 112: Number of glances collapsed across training conditions.

The Nav Entry task was associated with the largest mean number of glances (35.9), distantly followed by the Navigation Cancel (14.7) and Radio Manual Hard tasks (14.6). These results largely parallel task completion time and TEORT metrics (see section *Summary Comparison of Task Time and Glance Metrics* and Figure 118 & Figure 119).

Comparison to Study 1

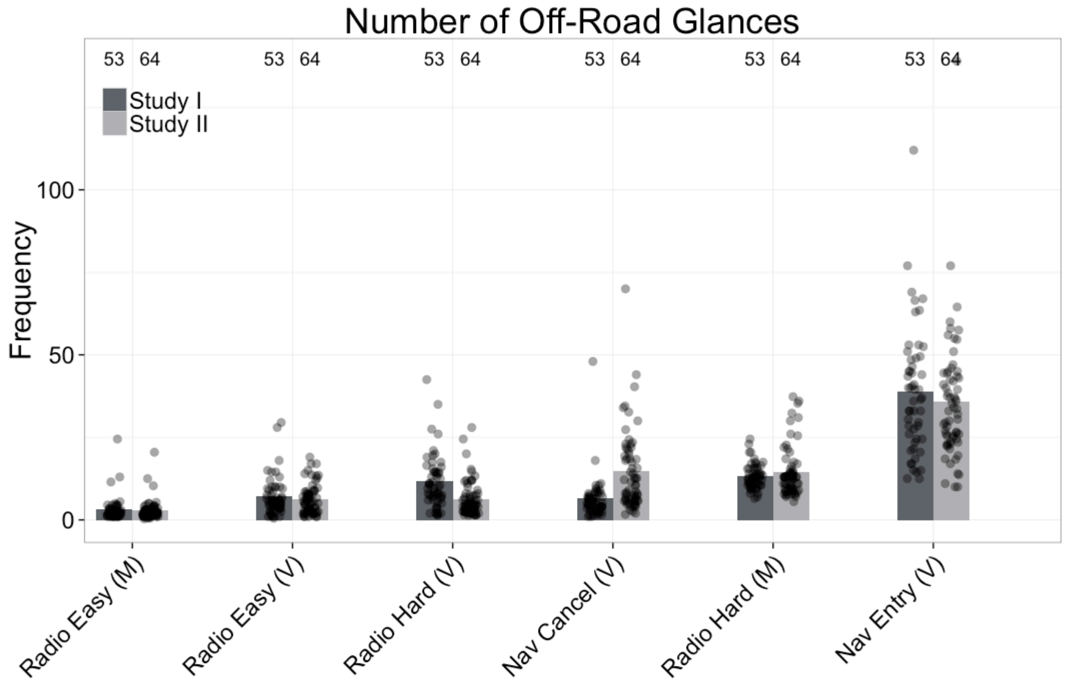


Figure 113: Comparison of number of glances in Study 1 and Study 2.

The increase in the number of TEORT glances in Study 2 for the Nav Cancel task is likely associated with the more difficult nature of this task in the second study as has been discussed in detail already. TEORT and total GTD times show consistently higher values in Study 2 as well. Similarly, as has been discussed already in the TEORT results section on *Comparison to Study 1*, the lower number of glances during the voice version of the Radio Hard task is understandable in terms of a reduction in the number of steps required to complete the task in Study 2. The same section discusses a possible age related explanation for the nominally lower number of EFOR glances in Study 2.

TEORT vs. GTD Comparison

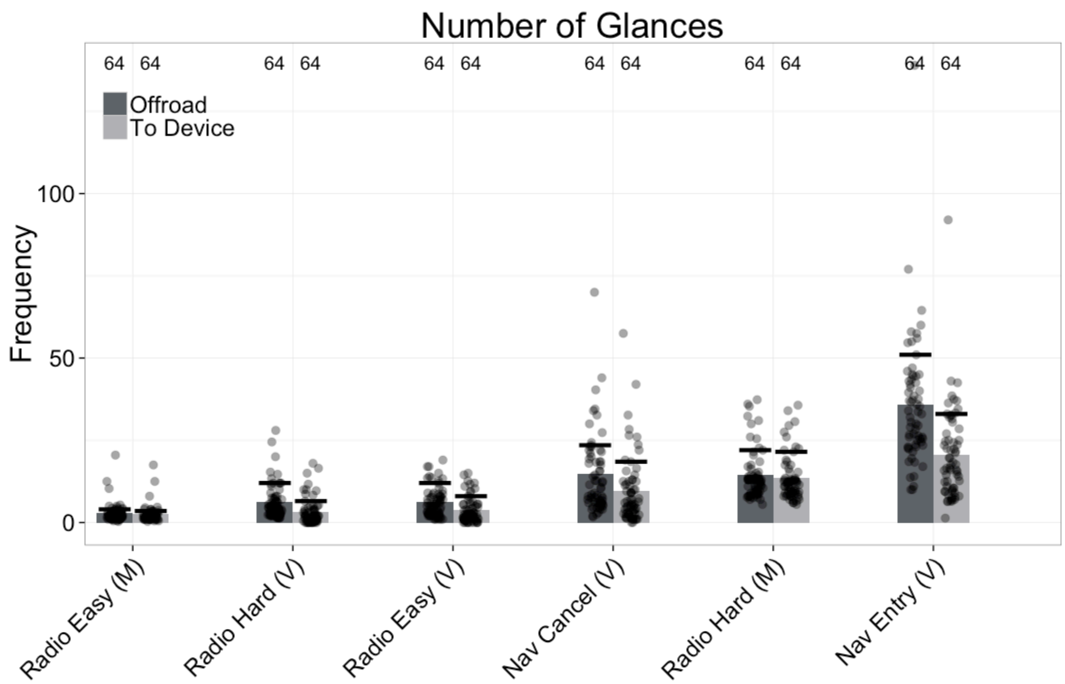


Figure 114: Comparison of number of glances during task period using TEORT and GTD metrics.

The same variation in the pattern of number of glances per tasks across the two studies follows that seen when comparing TEORT and GTD values for both studies (see Figure 99 and

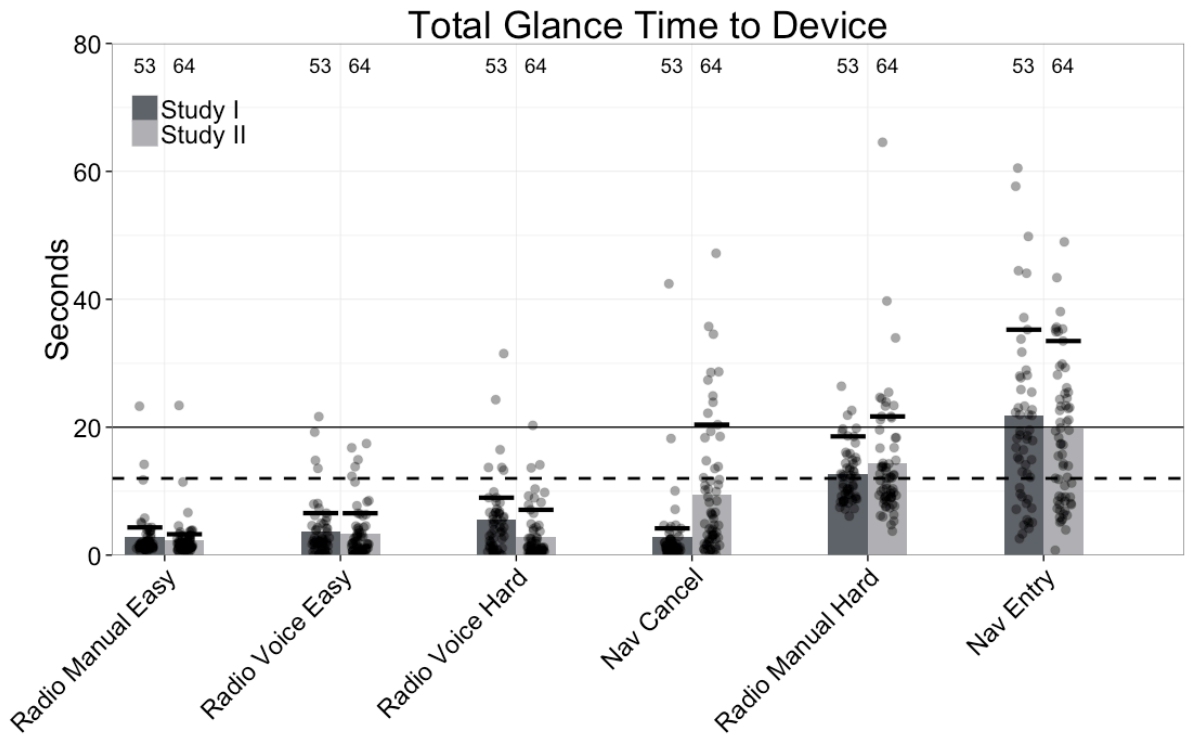


Figure 105). See comments in those sections.

Statistical Comparison of Selected Tasks

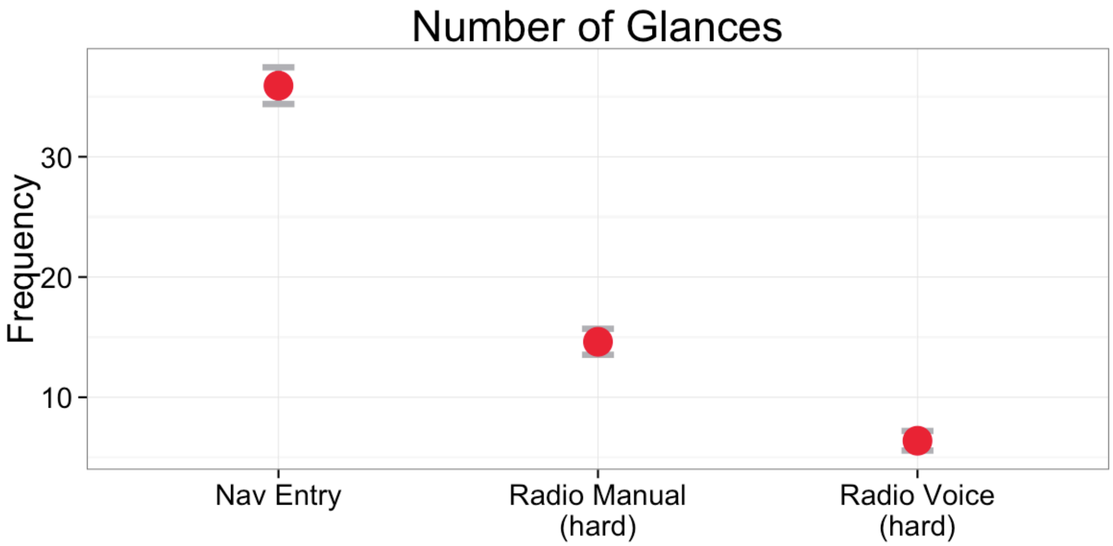


Figure 115: Number of glances across selected reference tasks.

The number of glances varied significantly across the three reference tasks ($X^2 = 105.1$, $p < .001$, Friedman test). Post-hoc testing reveals that all three tasks were significantly different from one another ($p < .001$ for all comparisons, Wilcoxon signed rank tests), again closely mirroring the TEORT metric.

Descriptive Statistics by Task with Age & Gender Breakdowns

Table 49: Mean and (SD) of number of glances by task, broken down by age group and gender.

Combined	**20-24**	**25-39**	**40-54**	**55+**	**(all)**
Nav Entry	34.19 (12.0)	32.48 (16.7)	34.92 (15.5)	42.07 (28.9)	35.91 (19.3)
Nav Cancel	17.17 (18.3)	10.62 (7.7)	13.06 (8.5)	17.74 (10.7)	14.65 (12.1)
Radio Manual Easy	2.39 (2.4)	2.31 (1.3)	3.14 (4.7)	3.72 (2.7)	2.89 (3.0)
Radio Manual Hard	12.85 (7.1)	13.69 (6.9)	13.58 (7.7)	18.33 (8.2)	14.61 (7.6)
Radio Voice Easy	4.06 (2.9)	6.54 (4.7)	6.47 (4.6)	8.20 (4.8)	6.32 (4.5)
Radio Voice Hard	6.78 (6.8)	6.73 (6.3)	5.38 (4.7)	6.65 (3.6)	6.38 (5.4)
Female	20-24	25-39	40-54	55+	(all)
Nav Entry	34.19 (14.7)	34.85 (19.6)	27.46 (12.6)	33.73 (13.1)	32.56 (14.8)
Nav Cancel	23.60 (23.6)	10.42 (8.2)	15.84 (9.5)	16.27 (12.1)	16.53 (14.7)
Radio Manual Easy	3.15 (3.2)	2.62 (1.4)	1.92 (0.6)	4.58 (3.6)	3.07 (2.6)
Radio Manual Hard	14.35 (9.6)	13.94 (4.5)	17.62 (9.0)	19.79 (9.9)	16.43 (8.5)
Radio Voice Easy	5.02 (3.8)	8.71 (5.0)	4.29 (2.7)	6.96 (4.1)	6.24 (4.2)
Radio Voice Hard	5.44 (3.4)	5.54 (4.2)	4.00 (2.3)	7.31 (3.9)	5.57 (3.6)
Male	20-24	25-39	40-54	55+	(all)
Nav Entry	34.19 (9.5)	30.10 (14.2)	42.38 (15.2)	50.42 (38.2)	39.27 (22.6)
Nav Cancel	10.73 (8.2)	10.83 (7.7)	10.27 (6.8)	19.21 (9.6)	12.76 (8.6)
Radio Manual Easy	1.62 (0.9)	2.00 (1.1)	4.35 (6.6)	2.85 (1.0)	2.71 (3.4)
Radio Manual Hard	11.35 (3.1)	13.44 (9.0)	9.54 (3.2)	16.88 (6.2)	12.80 (6.3)
Radio Voice Easy	3.10 (1.3)	4.38 (3.3)	8.65 (5.3)	9.44 (5.4)	6.39 (4.8)
Radio Voice Hard	8.12 (9.1)	7.92 (8.0)	6.75 (6.1)	5.98 (3.4)	7.19 (6.7)

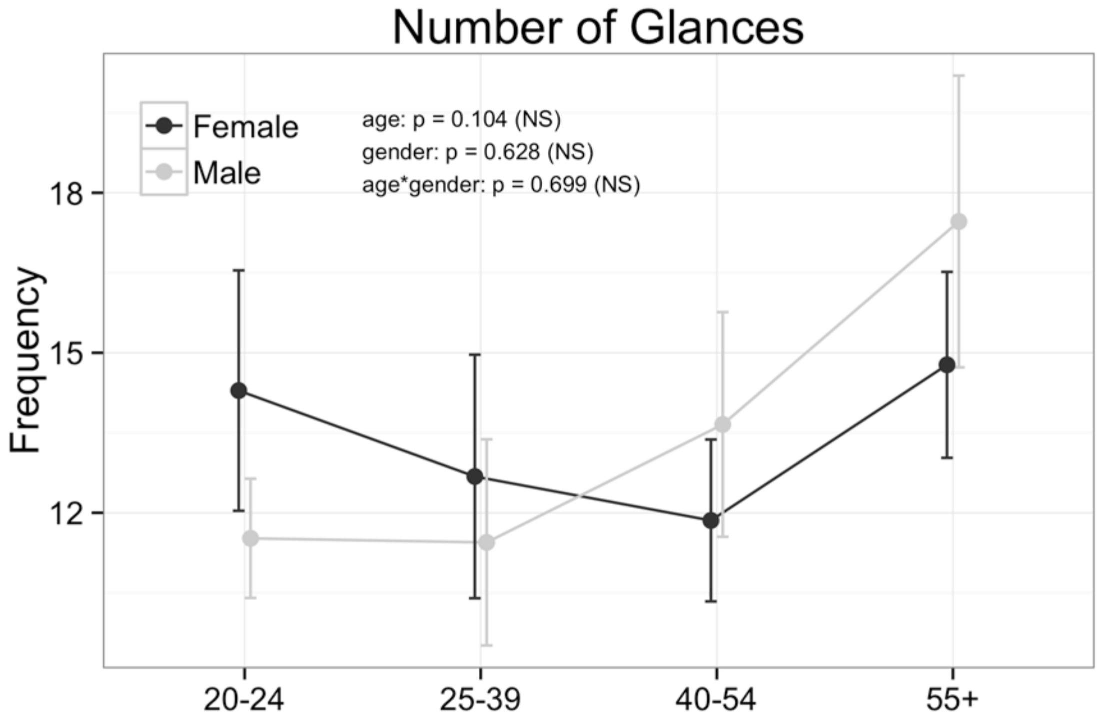


Figure 116: Number of glances across age groups and genders.

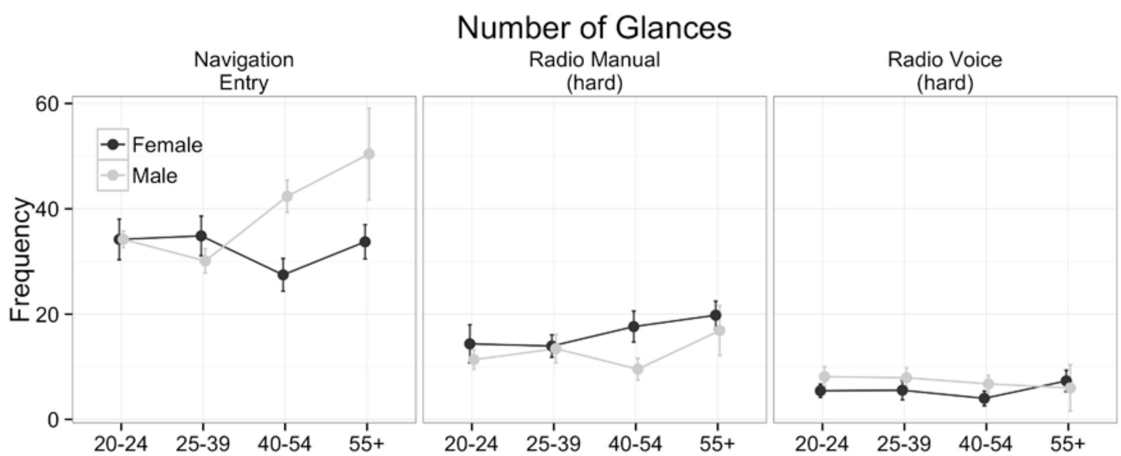


Figure 117: Number of glances by age and gender for the navigation entry task, radio manual (hard) task, and radio voice (hard) task.

Glance frequency did not vary significantly across ages or genders. However, there is a notable increase in glance frequency for older men when performing the Navigation Entry task.

Summary Comparison of Task Time and Glance Metrics

In discussing the importance of task duration and related measures in assessing the distraction potential of in-vehicle tasks, Burns, Harbluk, Foley and Angell (2010) highlight the close relationship between total task duration (completion time) and the total amount of time the eyes are directed off the roadway (TEORT) when assessing the demand associated with visual-manual interfaces. They also note that task duration is also likely relevant for in distraction for “more cognitive, auditory or speech-based tasks”, although they do not discuss this further in the paper. While Burns et al. provide a concise review of why TEORT is a good predictor of crash risk, they also emphasize that task completion time for visual-manual device interactions presents a much easier measure to collect and process and can be seen as very useful in this regard as a handy surrogate in early phases of the design and prototype testing process.

Figure 118 on the next page brings together the plots of task completion time, TEORT, and number of glances per task. It is apparent in the figures that there is a high degree of consistency in the ordering of the tasks across the three metrics, with the sole exception of the ordering of the Radio Manual Hard and the Nav Cancel tasks between task completion time and TEORT. Even that minor divergence makes some sense in that the longer task completion time for Nav Cancel may well be related to a number of participants pausing in an attempt recall the precise format of the voice command for which no visual support was directly available on the center console display (where only the route map was visible).

Perhaps the most meaningful divergence between task completion time and the TEORT and number of glance plots is in the relative magnitude of voice-involved radio “hard” tuning task to the manual radio “hard” tuning task. While the task completion times for both tasks are in the same general range, TEORT and number of glances are notably lower for the voice implementation. Thus, while the task duration is comparable across the voice-involved and the traditional visual-manual methods of completing the task, the total number of glances and total glance time is appreciably shorter. This is clearly in line with the design intent of providing a voice-command option for this task.

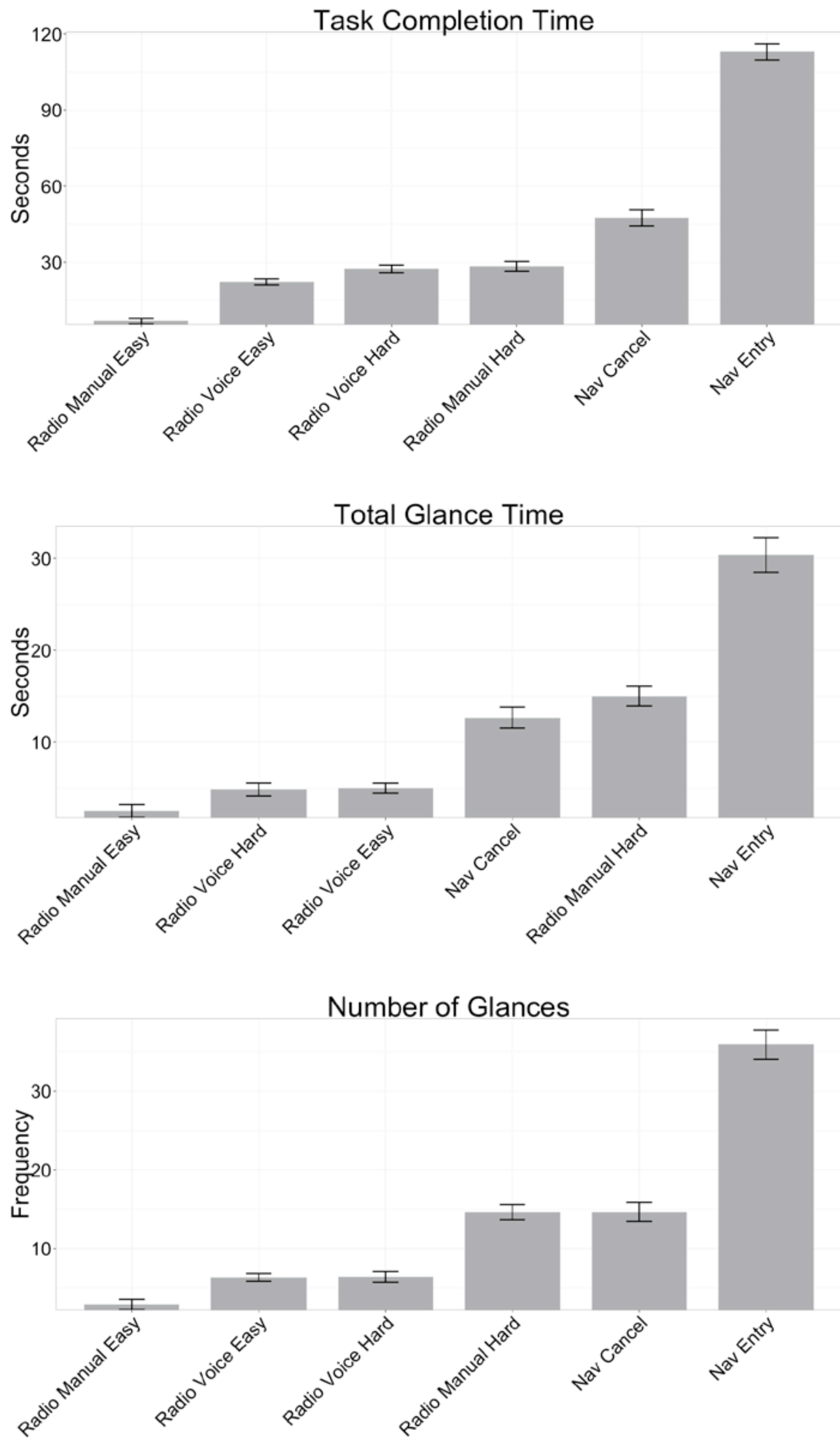


Figure 118: Combined presentation of Figure 18, Figure 97 showing TEORT, and Figure 111 to highlight the relationships between total task time, total time the driver's eyes are off the forward roadway, and the number of glances off the forward roadway. See previous page for discussion.

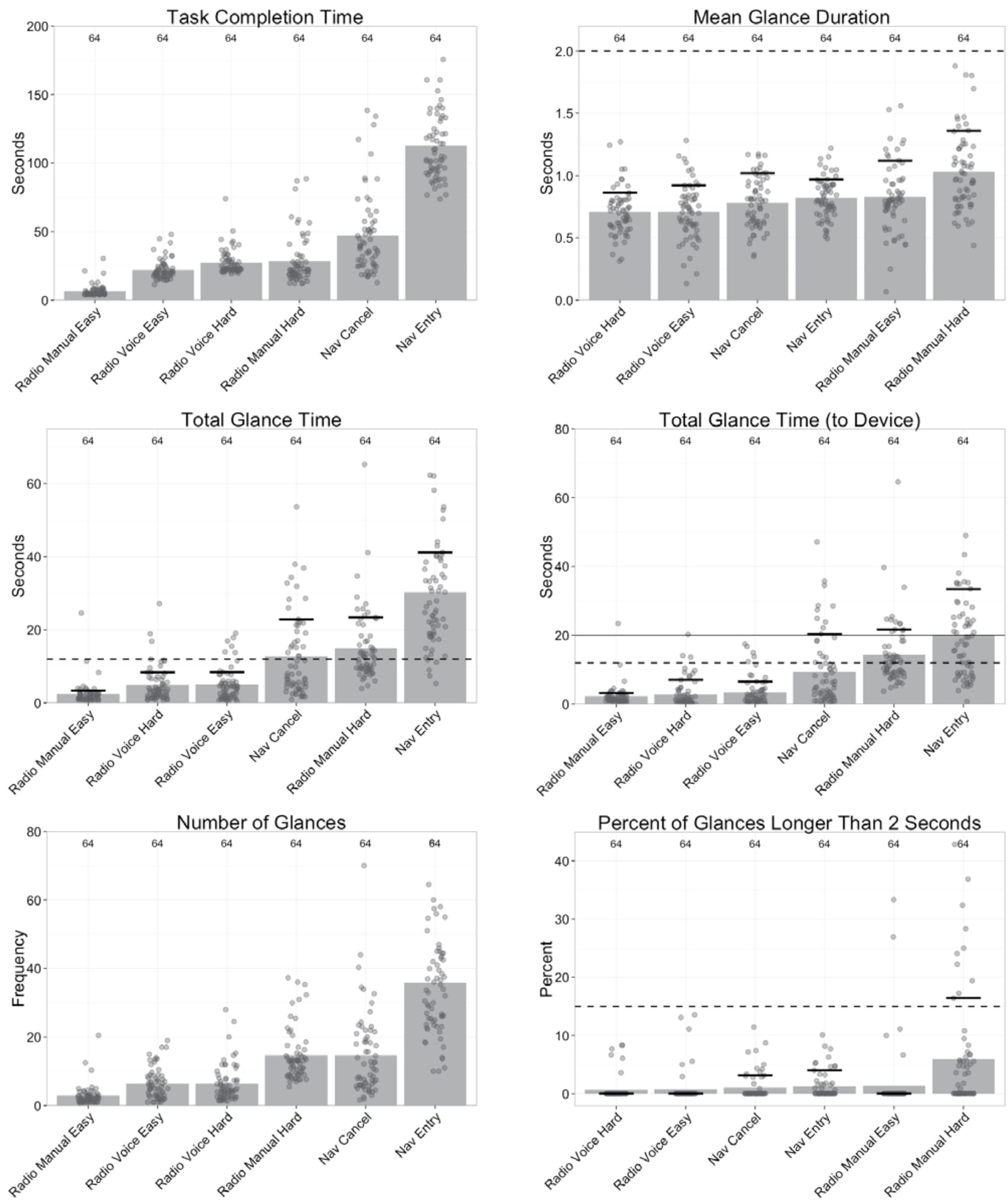


Figure 119: Combined presentation of findings for the distribution of individual participants across the DVI task types considering total task time and various eye glance metrics (TEORT unless listed as "to device"). See individual metric sections for discussion of each variable and larger figures.

Orienting Response

This assessment of participants' behavior toward the center stack graphic display was developed during the post-data collection review of participant behavior during Study 1. The rating attempts to characterize a behavior in which the participant appears to engage directly with the in-vehicle display, as if the voice-command interface were located within it. For example, the participant might begin speaking toward the display's location, lean towards it, change his posture, turn his body, or otherwise behave in a manner that suggests he has begun to prioritize interaction with the in-vehicle display. In Study 1, a single coder reviewed video of drivers during selected tasks and rated the level of orienting response displayed during various tasks. The detailed coding guide for orienting response behavior is given in the technical report for Study 1 (Reimer et al., 2013). In Study 1, these ratings suggested a markedly higher frequency of such orienting behavior in older participants (60-69 years) relative to younger participants (20-29 years), particularly in older females.

Although the results from Study 1 created substantial interest in the orienting response, full double coding and mediation of this behavior is highly resource intensive and was outside of the scope of the original project proposal for Study 2. Therefore, as in Study 1, a single coder manually assessed videos from the present study for indications of an orienting response, following the coding guide from Study 1. The figure on the next page should be considered only exploratory, as it is single coded and not mediated across coders. If only the youngest (20-24) and oldest (55+) age groups are considered, a trend suggesting higher orienting to the device again appears in the older vs. younger participants. However, the gender patterning seen in Study 1 does not appear to replicate. The age relationship also seems less clear if all age groups are considered in Study 2. Given that the individual cell sizes are half the size of those of the age groupings considered in Study 1 and that single coding was applied, the extent to which the findings of Study 1 should be considered to have or have not been replicated may be best deferred at the current time. Further evaluation of this behavior pattern may be worthy of additional evaluation in the future.

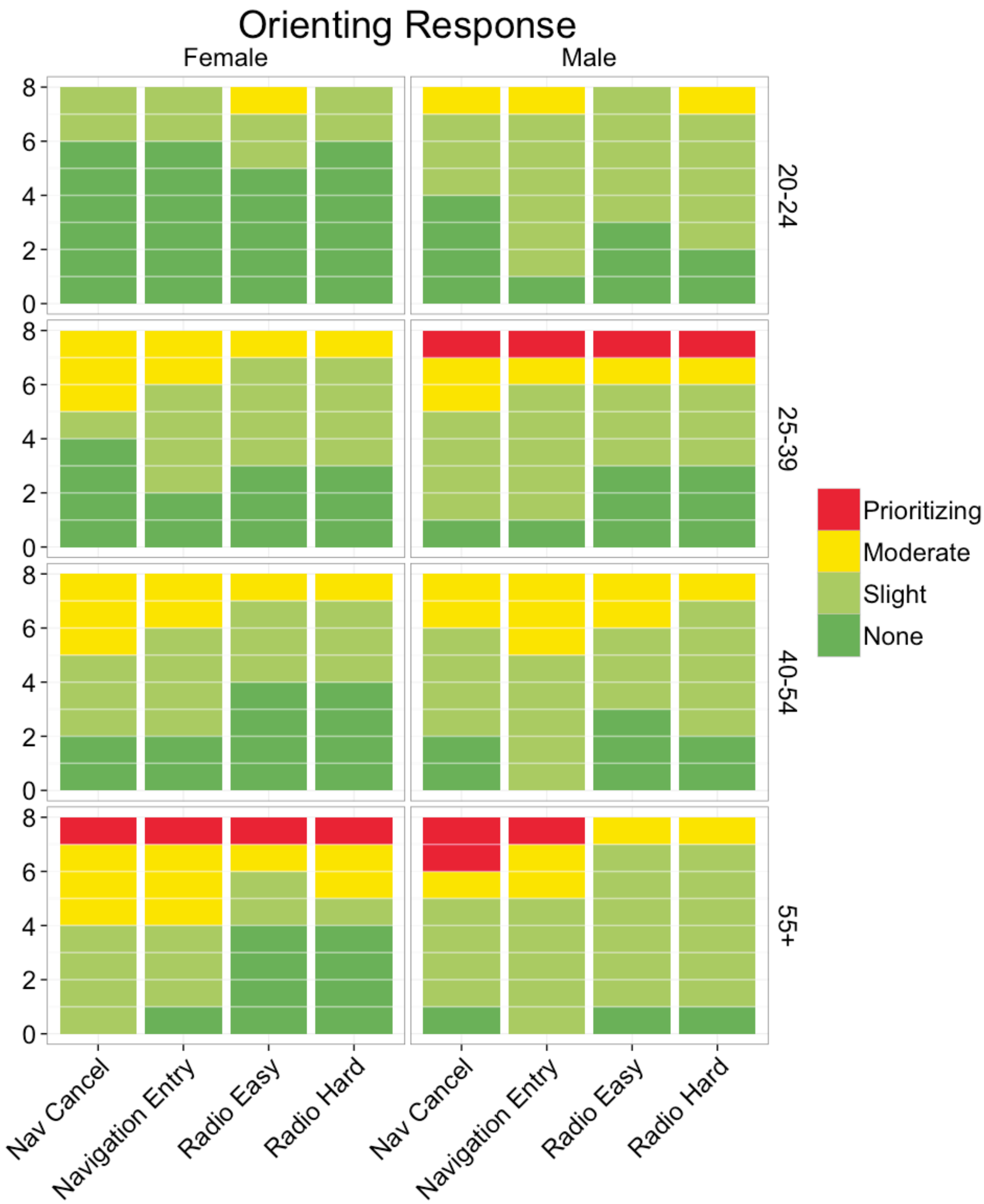


Figure 120: Exploratory evaluation of orienting response behavior based on a single coder, non-mediated assessment across tasks by age group and gender.

Task Performance

In our previous report, subjects were rated on their ability to complete the tasks of interest, and the extent of the assistance needed to do so (see previous report, "Task Completion Data"). For the present report, two coders independently rated task completion using the scale shown in Table 50. A third coder resolved any discrepant ratings.

Table 50: Task completion rating scale.

Rating	Definition
4	Participant completed task without error.
3	Participant completed task with 1 error.
2	Participant completed task with 2 or more errors.
1	Participant failed to complete task.

Note that this scale has been simplified from the one used in the previous report. In the previous report, the scale rated the degree to which the user needed help from the research assistant to complete the task. In the present study, the self-guided training group was never provided with help from the research assistant. To ensure that both groups would be comparable in these ratings, the scale was adapted to count errors instead of assistance. The previous report also showed that the vast majority of errors originated with the user, and not the system. Therefore, the scale was further simplified to count only errors from the user. Errors originating with the system (such as a failure to recognize a correct voice command) were ignored for the purpose of rating.

Task completion ratings are shown in Figure 121. The nature of these data makes a complete statistical analysis infeasible, but in particular, the difficulty of the Navigation Cancel task is worth examining in more detail. The self-guided training group experienced a high rate of failure on this task (16 out of 32 participants failed to complete it within a reasonable amount of time), and although the structured training group experienced a much lower failure rate (3 of 32), they still experienced a large number of errors. Only 14 out of 64 participants completed the task without error. A chi-squared test of independence suggests that structured training significantly reduced the severity of errors compared to the unstructured group (X^2 [df = 3] = 16.7, $p < .001$). These results draw a stark contrast with our previous report, in which 55 out of 60 subjects completed the task without error. This discrepancy is due to the fact that in the previous study, the command used to cancel the navigation route was explicitly stated during the pre-task instruction prompt.

The second-most difficult task, as judged by task performance ratings, was the Navigation Entry task. This task has a high rate of error, but a lower absolute rate of failure compared to the Navigation Cancel task. The Navigation Entry task requires a large number of steps to complete, which likely accounts for the high error rate. The Navigation Cancel task, on the other hand, is executed with a single command. If a participant failed to recall the exact command, the task could not be completed in the allotted time.

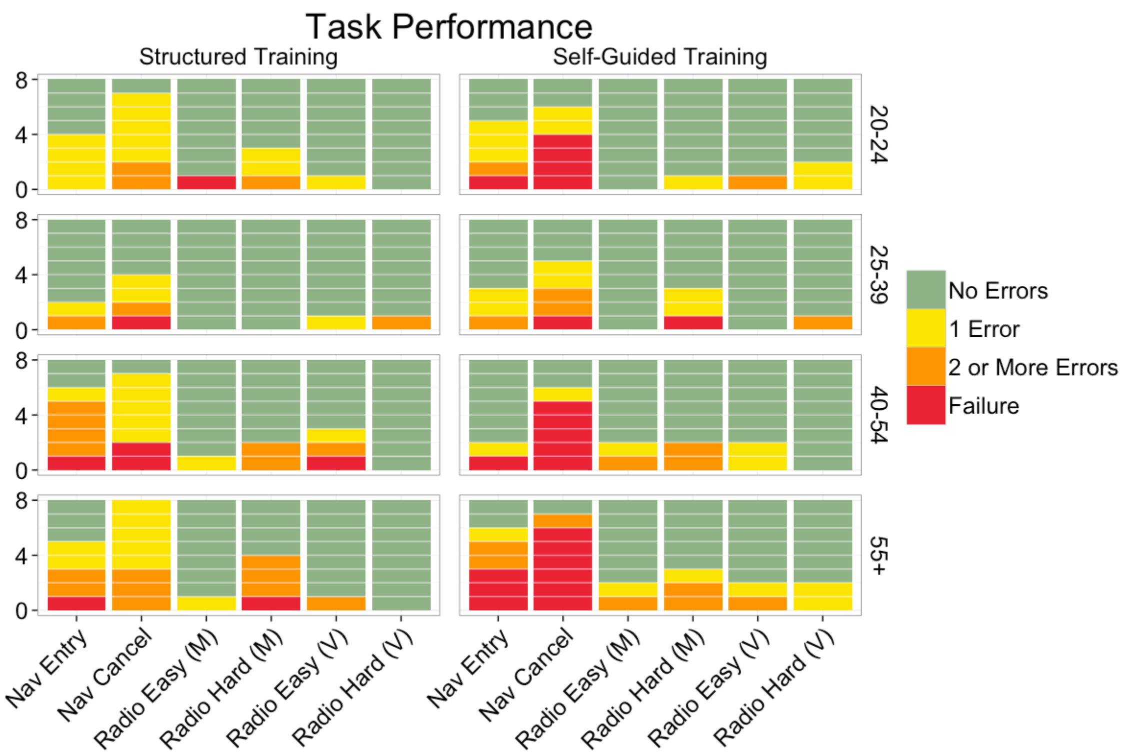


Figure 121: Task performance ratings organized by training group and age group. For each task, the worse of the two trial replications is shown.

Discussion

The plan to conduct a follow-on study to evaluate the significance of training method on a participant's interaction with a voice-based DVI was already in place while Study 1 was underway. A number of factors subsequently influenced the primary research goals and eventual design of the Study 2. Among these were the findings in Study 1 that the apparent visual demands associated with some of the voice-involved DVI tasks were markedly higher than initially anticipated. During this period, NHTSA (2012) released draft visual-manual driver distraction guidelines for in-vehicle electronic devices. While the NHTSA guidelines specified methodology for simulation based testing of DVIs, the findings for some of the voice-involved DVI tasks considered in Study1 suggested that, if applied to these tasks, the DVI may not meet the proposed guidelines. In our view, this strongly argued for the need to determine if the patterns of behavior observed in Study 1 would replicate. Further, Study 1 employed a sample of younger (20-29 years) and relatively older (60-69 years) drivers. The new NHTSA guidelines specified an equal distribution of participants across four age groups (18-24, 25-39, 40-54, and 55 and older). Study 2 was undertaken following this age distribution. Finally, the NHTSA guidelines specifically endorsed the concept of using multi-step manual radio tuning as basic reference task for establishing a maximum level of socially acceptable demand on the driver. The form of the "radio hard" task used in Study 1 was modified in Study 2 to explicitly conform to the specifications defined in the draft guidelines document (NHTSA, 2012).

Significance of Training Method

As already described, the initial focus of this second CSRC supported study was on evaluating the extent to which training on how to use a production voice-command system (likely unfamiliar to most participants) might impact a driver's willingness to try to use the system, ability to successfully use the system, and the level of workload and distraction associated with using the system. Given that users typically avoid reading user's manuals or experience limited, if any, guided instruction on the operation of a voice system at time of vehicle purchase, the impact of training is a highly relevant question.

Two training conditions were compared: *structured training* and *self-guided*. The first consisted of the same experimenter assisted, structured training protocol utilized in Study 1. This protocol took participants step-by-step through the most efficient method to complete a given task in the DVI's default mode. When using the voice-interface, this included training in shorthand techniques such as combining two verbal commands into a single command string to reduce steps. The intent of this structured approach as used in Study 1 was to ensure that the system under test was evaluated under optimal conditions for the default settings provided by the manufacturer. In the self-guided condition, participants were informed of the DVI tasks they would be asked to consider attempting while driving. Examples of the tasks and the form that they would be presented were provided in printed form so that it was clear exactly what would be encountered during the on-road evaluation. A defined period was independently allocated to each task type, during which the participant was encouraged to engage in self-directed experimentation with the DVI to discover how each task might be accomplished. The self-training condition was intended to assess many of the attributes of what a driver might encounter in a self-guided "driveway" training experience or when renting an unfamiliar vehicle. During the self-guided training period, participants had access to the DVI user's guide and were taken through the voice-calibration procedure which developed familiarity with the press-to-talk voice interface button, but no other basic experimenter support was provided.

We anticipated that a number of participants might find many of the voice-command tasks difficult to learn on their own and be unable to advance to actual on-road evaluation. While many of the self-guided participants struggled with certain aspects of selected tasks, taking participants through the voice calibration procedure in the test vehicle appeared to provide enough familiarity with the basics of the voice-command interface to give the self-guided participants the basic knowledge of how to start interacting with the system. Various auditory and visual support prompts provided as part of the DVI did in fact provide sufficient guidance for the majority of participants in the parking lot exploration experience to learn how to carry-out the specified tasks. One exception to this statement is the challenge that many participants in both training conditions found in canceling a navigation route request. While the majority of participants in the structured training group had some difficulty with the Nav Cancel task, all but 3 were able to successfully complete the task while driving. In contrast, half of the self-guided training group (16 participants) were unable to successfully cancel the route request. Recent experience by our group with other DVIs indicates that this apparent challenge in discovering and remembering the exact command syntax for canceling navigation is not unique to the DVI under study here.

Based on what we observed in the self-trained group, OEMs may wish to keep in mind the possible value that some exposure to the system's logic / vocabulary during a interaction like a voice calibration procedure (often employed to improve a system's recognition performance) may provide in assisting drivers in developing some concentrated experience with the command structure, pacing style, etc. that is required to interact with the voice-interface in an optimal way. A strategy such as this may be more likely to get a consumer engaged in an "interactive educational experience" to explore the basic features of advanced or otherwise unfamiliar DVIs than solely relying on the user's manual. In essence, suggesting to drivers that they take the system though a calibration procedure to improve its function may be a means to increase the likelihood that more drivers will in fact take a short amount of time to actually become more educated in how to interact successfully with the system. It is worth noting along these lines that none of the participants in the self-guided training group was observed to pick-up and refer to the user's guide even though it was placed on the front passenger seat next to them and they were explicitly told that it was there and available for reference. (Such a suggestion may well need to be presented on a temporary tag or removable sticker in addition to being provided in a user's guide as many drivers may never open the user's guide as we observed here. Such an approach might also be recommended as a standard component of the dealer training provided as part of a purchaser's introduction to the vehicle.)

Contrary to our expectations, no statistically significant main effect differences were observed across task types for the two training groups on any of the primary outcome measures (see *Summary Findings Comparing Training Conditions* and Table 4). Subjective impressions of the two experimenters who conducted the DVI training and on-road portions of the study suggest that the learning styles of individual participants varied greatly; it is possible that the individual differences of participants that found each type of exposure beneficial or hindering was such that no net effects across training groups was observed.

Because there were no statistically significant differences across training conditions the data collected in both training groups could be combined into a single large sample of 64 participants for purposes of assessing overall driver behavior. This was fortuitous in that it allowed for a fairly sensitive evaluation of the extent to which key findings from our first voice-interface study (Study 1 – Reimer et al., 2013) did or did not replicate in a second sample.

Replication of Findings from Study 1

In Study 2 we deliberately reduced the number of DVI tasks presented to participants to more closely approximate the maximum number of new task experiences a driver might attempt during a single long drive. Specifically, voice-based song selection from music stored on an external device and the phone dialing tasks included in Study 1 were not included in Study 2. The tasks that were evaluated consisted of both traditional visual-manual tuning and voice-command tuning of the radio, voice-command involved entry and cancelation of a full street address with the navigation system, and multiple levels of the auditory presentation – verbal response n-back surrogate cognitive load task. While numerous observations can be made in the detailed analysis of individual tasks (see individual measure sections in *Results*), the overarching summary finding is that the basic pattern of results considering self-reported workload, physiological arousal, driving performance metrics, and glance metrics seen in Study 1 largely replicate in the current study.

Where differences did appear, they seem reasonable given known differences between Study 1 and 2, such as the change in the Nav Cancel task where participants had to recall the specific command phrase to use to cancel route guidance instead of explicitly being told how to cancel the route as part of the task prompt. As a consequence, the task was rated as more difficult, took longer, was associated with greater physiological arousal, etc. in Study 2. In the same vein, the voice-command version of the radio "hard" task in Study 2 was rated nominally lower in workload, lower in skin conductance, lower in TEORT, and involved a lower number of glances than was the case in Study 1. This is logically attributable to the radio "hard" task being undertaken with the radio already in the "on" state in this second study. This resulted in one less voice command being required to accomplish this task than was the case in Study 1.

In brief, when using the NHTSA specified age distribution and format for the manual radio turning reference task (NHTSA, 2012), the following findings from Study 1 again were seen in Study 2:

- Voice recognition was found to be fairly robust with only 3 out of more than 80 participants unable to participate due to voice recognition issues.
- Apparent cognitive processing demand / workload (as assessed through heart rate and SCL) the DVI tasks studied fell below the level of the 1-back cognitive reference task.

- For the radio tuning reference task (Radio Manual Hard), the voice-command method was associated with lower workload (self-report heart rate, SCL), lower mean glance durations, a markedly lower percentage of long duration glances, and significantly lower total glance time than the visual-manual interface.
- Voice-command involved entry of a full destination address into the navigation system was associated with TEORT significantly above the acceptable criterion defined by NHTSA for visual-manual DVIs if the criterion was applied to this task.

Some expansion on these points follows.

Using Voice to Tune the Radio

For what has traditionally been a fairly intensive visual-manual secondary task in the vehicle, using the voice-command option in the DVI under study to obtain the same goal state as what is required in the NHTSA version of the radio reference task does appear to have clear advantages. The manual form of the reference task interaction involved pressing a mode button, making a touch-screen button "press" to switch between AM and FM bands, and rotating a manual tuning knob through multiple rotations to tune to the specified station. In contrast, the voice interface required the single button press of the voice system activation button, saying the frequency of the desired station (i.e. "100.7"), and saying "yes" to confirm the action. As already noted, participants rated this method as involving lower workload and it was associated with lower heart rate and SCL values, lower mean glance durations, a markedly lower percentage of long duration glances, and significantly lower total glance time. (See *Summary Findings Comparing Manual & Voice-Based Radio "Hard" Tuning,* in particular Table 5 and Table 6, for details on consistency between the two studies and significance levels for comparisons between the two operational methods.) These findings support the position that an appropriately designed voice-command interface can reduce the overall workload associated with this task. Further, the interaction with the DVI considered here did, for this task, fulfill the often stated goal of keeping the driver's eyes on the roadway to a much greater extent than what is observed tuning the radio manually.

As a method for replacing the visual-manual station selection using a station pre-set button, voice-command did not necessarily offer an overall advantage. Task completion time was longer and TEORT and number of glances higher. On the other hand, using NHTSA's TEORT version of the percentage of single long duration (>2s) glances metric, 6 out of 53 participants in Study 1 (4 in their 60s) and 2 out 64 in Study 2 exceeded the 15% criterion when manually selecting a pre-set button; none of the participants in either study exceeded the NHTSA criterion using the voice-command method for selecting a pre-set. Both methods were rated comparably by participants on self-reported workload. This may suggest some advantage in maintaining availability of single manual button interactions for frequent tasks for drivers with relatively good visual-manual coordination, while adding the voice-command option for drivers who may show some compromise in this area due to age or other factors.

In summary, a comparison of the visual demand metrics for the manual and voice-command based methods of accomplishing the radio hard task clearly indicate that an appropriately implemented voice-interface can reduce the visual engagement associated with accomplishing a given task solely through visual-manual interaction. However, this does not mean that a voice-involved interface will necessarily be free of visual demand. This is particularly evident when considering use of a voice-command interface to enter a destination address into a navigation system.

Using a Voice Interface for Destination Entry

Perhaps the most marked finding in Study 1 was the high TEORT value observed during the full address destination entry into the navigation system. In a sample of drivers meeting the NHTSA age distribution requirements (Study 2), setting a destination through discrete entry of *city*, *street name*, and *number* resulted in a mean TEORT of over 30 seconds – well in excess of the 12 second threshold established by NHTSA if it were applied to this DVI. Evaluated as a percentage of drivers within the threshold, only 7.8% of the sample fell within the 12 second threshold – well below the 85% criteria. Applying The Alliance GTD metric, total glance time drops to a mean for the sample of 20 seconds (SD 15.7s); however, this still results in only 58% of the sample falling within this threshold (see Figure 104). In addition, it should be noted that this calculation does not include glance time to the push-to-talk button.

On the other hand, as was the case in Study 1, the results for Study 2 show that the glance behavior during the destination entry task fell well within NHTSA's mean glance duration and percentage of single long (>2s) duration glance criteria. Examination of the destination entry task makes it apparent that it involves many more steps than the manual radio tuning reference task and it extends over a much longer period of time (M 113s, SD 31s vs. M 28s, SD 18s). While there is understandable concern over any activity that draws the eyes away from the forward roadway, the data from the destination entry task highlights the question of how total glace duration metrics should be considered in the context of longer duration tasks involving numerous discrete glances that otherwise fall within guidelines for glance characteristics (mean glance time and percentage of long duration glances). Both the original Alliance and the more recent NHTSA total glance duration metrics were developed in the context of using manual radio tuning as a reference point for a socially acceptable maximum level of visual-manual demand. If a new-generation task consists of more discrete steps than classical manual radio tuning and spaces them out over a longer interval than is typically devoted to manual radio tuning reference task, is there a point at which a somewhat longer total glance duration again might begin to be considered within a reasonable range from a total visual demand and safety risk perspective?

At the same time, it is important to consider if there is a point at which an overtly low to moderately demanding task becomes problematic due to length of engagement? Does the ability to self-pace a task effectively fully compensate for extending the time required to complete the task? Burns et al. (2010) provide a useful discussion of some of the complexity in considering the relationship between task time and the amount of time a driver is looking off the roadway. See *Summary Comparison of Task Time and Glance Metrics* in this report for a concise view of the relationship between these variables across the DVI tasks considered.

We noted in our reporting on Study 1 that much of the glance behavior observed during voice tasks was associated with looking at a console display screen to view options presented by the system, such as available commands or to make a selection from a list if the system identified multiple options for street names during address entry. The number of support and confirmatory steps was much greater for the navigation entry task than for any of the other tasks. Consequently, the total glance time metrics were directly impacted by the number of glances in the task (see Figure 118 and Figure 119). While inclusion of these visual support displays seems in one sense to be directly counter to the intent of a voice-command interface to reduce the extent to which drivers take their eyes off the roadway, it seems likely that these types of support displays are generally employed to reduce the amount of cognitive load that would be placed on the driver by having to remember specific command phrases or needing to listen to and hold in memory an extended list of destination selection options. It is worth pointing out in this context that the physiological measures monitored during the destination entry tasks showed arousal levels markedly lower than what was observed during the

1-back level of surrogate cognitive demand task (Figure 25 and Figure 32). This suggests that working memory demands were kept at a reasonable level during this multi-modal approach to supporting destination entry. These findings highlight the human factors design challenge to find a balance across visual, manual, auditory, and associated cognitive demands to optimally support drivers of differing capabilities and interface interaction preferences.

Additional Comments

We noted in the reporting for Study 1 (Reimer et al., 2013), the NHTSA visual-manual distraction guidelines (NHTSA, 2013) explicitly stated that they "are currently not applicable to the auditory-vocal portions of human-machine interfaces of electronic devices." NHTSA has since released a supplementary report, *Explanatory Material About the Definition of a Task Used in NHTSA's Driver Distraction Guidelines, And Task Examples* (Angell, Perez & Garrott, 2013), that states "Some task interactions involve mixed-mode interactions: a mixture of both auditory-vocal and visual-manual interactions. Because such tasks do involve some visual-manual interaction, it is appropriate that the visual-manual components of these tasks meet the proposed Phase 1 NHTSA Distraction Guidelines" (p. 32). However, it is not entirely clear from these statements whether the visual-manual and auditory-vocal components of tasks such as the multi-step, full destination entry task considered in these studies should somehow be considered. It seems inherent in the second statement that glances to the display when it presents the driver with a visual listing of destination options should be included in the calculation of glance metrics. On the other hand, if the driver looks at the display screen to confirm if the system correctly understood a verbal command at another point in the destination entry task, should this be excluded since it might be considered as associated with an auditory-vocal portion of the task? An alternate approach is to assume that this glance away from the roadway was associated with the overall DVI interaction and should be counted, which was done in the EORT analyses in this report. Perhaps an even grayer area is the question of the inclusion of other glances, such as mirror inspections, that may occur multiple times during extended mixed-mode interactions. If visual engagement is considered over the full length of the task, should the broader temporal components of multi-modal tasks be considered in future guideline work such as the development of Phase III distraction guidelines? As already detailed, the mixed-mode destination entry task extended over a time period that averaged four times the duration of the manual radio tuning reference task.

The gender and age distribution characteristics of the participants in this on-road study fully comply with NHTSA guidelines for DVI evaluation. As detailed in the body of this report, statistically significant main effects of gender appeared with males showing longer mean single glance durations and a higher percentage of long duration (> 2 second) glances. Age showed a statistically significant main effect on TEORT, with eyes-off-road time being higher in the older age groups. Similar trends were seen in mean single glance duration and percentage of long duration glances. Thus, not unexpectedly, age and gender factors can influence assessment measures. NHTSA's recommendation that evaluation samples be balanced by age and gender, and that the age distribution include a representative sample of older drivers is quite relevant.

As we emphasized in our earlier reports on Study 1 (Reimer & Mehler, 2013; Reimer et al, 2013), there is no a priori reason to assume that the visual demands observed here are unique to the specific voice-command interface evaluated in this study. Other systems employing similar design characteristics are likely to demonstrate similar visual demand involvement. Initial reviews of systems produced by other manufacturers carried out by our group and other researchers (e.g. Reagan & Kidd, 2013) also suggest a high degree of multi-modal input and output characteristics in their voice-involved DVIs. However, future work will need to formally assess the degree to which the results reported here generalize to other production systems.

Conclusions & Limitations

Summarizing the key observations across Study 1 and Study 2, it is apparent in the assessment of the voice-command interface in these studies that the attentional draws in modern DVI's can be highly multi-modal. Depending upon a given design, newer DVIs can involve various combinations of demand (visual, manual, auditory, vocal, haptic, cognitive, etc.). Simply including voice in a DVI does not preclude the need to consider possible visual demand characteristics that may be present. Visual as well as other potential demand sources need to be included in the assessment of voice-command interfaces. Based on the measures collected here (self-reported workload, peripheral physiological arousal, basic driving metrics, and comparative surrogate cognitive tasks), cognitive demand may be relatively moderate in appropriately designed voice-command interactions. Cognitive demand becomes more apparent when drivers have difficulty completing activities ("song-fail" task in Study 1 or have difficulty recalling command syntax such as "navigation cancel route" in Study 2).

Observations from the navigation cancel task suggest that ensuring a user has an established understanding of a system's syntax is necessary when evaluating the fundamental cognitive demand of an interaction. This may need to be considered independently from the cognitive resources invested by a driver who finds the interaction non-intuitive. Such a situation may arise when drivers have developed models of interaction (command syntaxes) using other voice enabled technologies (smartphones, other vehicles, etc.). This latter finding is a critical consideration when evaluating the demands that real drivers may experience with systems that they have become familiar with as opposed to systems that they are encountering for the first time or have overall low exposure.

As a potential limitation, it should be noted that the measures of cognitive demand in the present studies were not exhaustive. Further, it is conceivable that other voice-command implementations might be associated with overtly higher levels of cognitive demand. For example, while some of the support displays provided as part of the destination entry task introduced visual demand, allowing the driver to glance at a list of street name options might well be less cognitively demanding than presenting an auditory list of numbered options that a driver has to hold in memory prior to making a selection. Again, evaluations of other voice-command systems are in order to determine the generalizability of the findings of the current studies. It is also unknown how this field data correlates with findings that would be obtained under the simulation procedures specified in the Phase I visual-manual guidelines.

Next Steps

Our work will continue to evaluate implementations of voice-command systems. As this is a rapidly improving technology, the primary goal in evaluating additional systems will be to develop a broader set of recommendations based on empirical data that DVI designers will find useful in further optimizing the demand that multi-modal interfaces place on drivers.

Acknowledgements

Acknowledgement is extended to the Toyota Collaborative Safety Research Center (CSRC) which provided the primary funding for this project. In addition, we are particularly grateful for the valuable, constructive comments that James Foley and Kazutoshi Ebe of CSRC provided during the development of the study.

Supplemental / matching support was provided US DOT's Region I New England University Transportation Center at MIT. The vehicle under test was purchased through funding from Ford Motor Company for an earlier project assessing the Ford Active Park Assist™ feature (Reimer, Mehler, & Coughlin, 2010) and support for originally instrumenting the vehicle was supplied by The Santos Family Foundation. The current project was conducted without consultation or involvement of Ford Motor Company.

This work would not have been possible without the support of AgeLab staff and visiting scholars including: Erin McKissick, Enrique Abdon Garcia Perez, Adrian Rumpold, Thomas Manhardt, Brahmi Pugh, Martin Lavalliere and Brendan Drischler in the development and refinement of data analysis and extraction tools, and exhaustive reduction and coding of eye glance and other data.

The interpretive aspects of this report reflect the views of the authors, who are also responsible for the factualness and accuracy of the information presented herein.

References

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., et al. (2006). Driver Workload Metrics Task 2 Final Report. Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration.

Angell, L., Perez, M., & Garrott, W. R. (2013, November). Explanatory material about the definition of a task used in NHTSA's driver distraction guidelines, and task examples. (Report No. DOT HS 811 858). Washington, DC: National Highway Traffic Safety Administration.

Burns, P., Harbluk, J., Foley, J., & 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 of Automotive User Interfaces and Interactive Vehicle Applications (Automotive UI 2010),* November 11-12, Pittsburgh, PA, USA.

Barón, A., & Green, P. (2006). Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review. Ann Arbor, MI: The University of Michigan Transportation Research Institute (UMTRI).

Carter, C., & Graham, R. (2000). Experimental comparison of manual and voice controls for the operation of in-vehicle systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 44, No. 20, pp. 3-286). SAGE Publications.

Chiang, D. P., Brooks, A. M., & Weir, D. H. (2005). Comparison of Visual-Manual and Voice Interaction with Contemporary Navigation System HMIs. *SAE Technical Paper 2005-01-0433*.

Driver Focus-Telematics Working Group. (2006). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems, Version 2.0: Alliance of Automotive Manufacturers.

Ei-Wen Lo, V.E-W. & Green, P. A. (2013). Development and Evaluation of Automotive Speech Interfaces: Useful Information from the Human Factors and the Related Literature. *International Journal of Vehicular Technology, 2013*. http://dx.doi.org/10.1155/2013/924170

Harbluk, J., Burns, P. C., Lochner, M., & Trbovich, P. L. (2007). Using the lane-change test (LCT) to assess distraction: Tests of visual-manual and speech-based operation of navigation system interfaces. Paper presented at the 4th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Stevenson, WA.

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis and Prevention, 38(1),* 185-191.

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Human Factors, 54(3),* 396-412.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record (2138),* 6-12.

Liu, Y., Schreiner, C. S., & Dinges, D. T. (1999). Development of human factors guidelines for advanced traveler information systems (ATIS) and commercial vehicle operations (CVO): Human Factors Evaluation of the Effectiveness of Multi-Modality Displays in Advanced Traveler Information Systems (FHWA-RD-96-150). McLean, VA: Office of Safety and Traffic Operations R&D Federal Highway Administration.

Lerner, N., Singer, J., & Huey, R. (2008). Driver strategies for engaging in distracting tasks using in-vehicle technologies (Report No. HS DOT 810 919). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA)

National Highway Traffic Safety Administration. (2012). *(Proposed Guidelines)* Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).

National Highway Traffic Safety Administration. (2013). *(Issued Guidelines)* Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices

(Docket No. NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).

Östlund, J., Nilsson, L., Carsten, O., Merat, M., Jamson, H., Jamson, S., et al. (2004). Human machine interface and the safety of traffic in europe (HASTE) project deliverables 2: HMI and safety-related driver performance.

Östlund, J., Peters, B., Thorslund, B., Engström, J., Markkula, G., Keinath, A., et al. (2005). Adaptive Integrated Driver-Vehicle Interface (AIDE): Driving performance assessment - methods and metrics. (Report No. IST-1-507674-IP). Gothenburg, Sweden: Information Society Technologies (IST) Programme.

Owens, J. M., McLaughlin, S. B., & Sudweeks, J. (2010). On-road comparison of driving performance measures when using handheld and voice-control interfaces for mobile phones and portable music players. *SAE International Journal of Passenger Cars – Mechanical Systems, 3*(1), 734-743.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Reagan, I.J. & Kidd, D.G. (2013). Using hierarchical task analysis to compare four vehicle manufacturers' infotainment systems. *Proceedings of the Human Factors and Ergonomic Society 57th Annual Meeting*, pp. 1495-1499.

Reimer, B. (2009). Cognitive task complexity and the impact on drivers' visual tunneling. *Proceedings of the Transportation Research Board of The National Academies*, Washington, DC, 13-19.

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics, 54(10),* 932-942.

Reimer, B. & Mehler, B. (2013). The effects of a production level "voice-command" interface on driver behavior: summary findings on reported workload, physiology, visual attention, and driving performance. MIT AgeLab White Paper No. 2013-18A. Massachusetts Institute of Technology, Cambridge, MA.

Reimer, B., Mehler, B., & Coughlin, J. F. (2010). An evaluation of driver reactions to new vehicle parking assist technologies developed to reduce driver stress (MIT AgeLab White Paper). Cambridge, MA: Massachusetts Institute of Technology.

Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. F. (2013). The Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Reported Workload, Physiology, Visual Attention, and Driving Performance (Technical Report 2013-17A). Cambridge, MA: MIT AgeLab.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2012). A field study on the impact of variations in short-term memory demands on drivers' visual attention and driving performance across three age groups. *Human Factors, 54(3)*, 454-468.

Shutko, J., Mayer, K., Laansoo, E., & Tijerina, L. (2009). Driver workload effects of cell phone, music player, and text messaging tasks with the Ford SYNC voice interface versus handheld visual-manual interfaces *SAE Technical Paper 2009-01-0786*. doi: 10.4271/2009-01-0786

Son, J., Mehler, B., Lee, T., Park, Y., Coughlin, J. F., & Reimer, B. (2011). Impact of cognitive workload on physiological arousal and performance in younger and older drivers. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Lake Tahoe, CA, 87-94.

Appendix A: Self-Reported Workload Materials

As noted in the main body of this report, self-reported workload ratings followed the same general form as that used in Study 1 and background on the approach is provided in Appendix I of that report.

The rating form used in this study varied somewhat from that of Study 1due to some differences in the set of tasks studied. Therefore, the rating form used is reproduced on the following pages for reference purposes. One other variation was present in the procedure from that of Study 1. In Study 1, tasks presented during the first half of the drive were rated at the mid-point rest stop and the tasks presented during the second half of the drive were rated upon the completion of the study back at MIT. In Study 2, the total number of tasks was fewer in number (no *song selection* or *phone dialing*), so all "training" experience was provided in the parking lot at MIT as opposed to being divided between the MIT parking lot and a mid-point stop. Since the mid-point stop from Study 1 was eliminated, self-report ratings of all tasks were done at the conclusion of the drive back at MIT.

Instructions

The following instructions on how to fill out the workload rating scales were given to participants to read by a experimenter. Participants were given an opportunity to ask any questions that they had. The page following this one reproduces the paper and pencil rating form that was given to participants at the mid-point rest period to rate the tasks completed during Part I of the drive, and again at the completion of the experiment to rate the remaining tasks. The form was given on a larger, legal size sheet (11x17 inches) and clipboard; it is reduced in size here. The standard sizing supports a 10cm long visual scale using the numbers 0 to 10 but with the option to circle midpoint values corresponding to a potential 21 scalar values.

These instructions are identical to those used in Study 1.

How to Fill-Out the Workload Rating Scales

One of the goals of this study is to understand how much effort is involved doing different tasks while driving. At the mid-point rest stop and at the end of the drive, we will ask you to fill out a short form to rate how much workload was associated with each type of task using a 0 to 10 scale where 0 = very low workload and 10 = very high workload.

By workload, we mean the demand or effort that was associated with trying to do a task while continuing to drive safely. Workload is experienced in different ways by different people. Workload could mean the amount of mental effort that was involved in the task, the amount of attention that was required, how physically difficult it was to complete, how much time pressure was involved, or how frustrating the task was. You can decide for yourself how to define workload as it relates to doing these tasks while also continuing to drive safely.

The instructions shown below will appear on the form:

Please circle a point along each scale that best corresponds to how much **workload** you felt was involved in trying to do each task. Workload is best defined by the person doing the task and may involve *mental effort*, the amount of *attention required*, *physical effort*, *time pressure*, *distraction* or *frustration* associated with trying to do the task while continuing to drive safely.

Example of Format that will be used to rate the workload associated with a task:

Task XXX

● ●

Low 0 1 2 3 4 5 6 7 8 9 10 High

Please let the experimenter know if you have any questions about how to fill-out this form.

Thank you.

Example Rating Sheet

Workload Rating – 2013a

Please circle a point along each scale that best corresponds to how much **workload** you felt was involved in trying to do each task. Workload is best defined by the person doing the task and may involve *mental effort*, the amount of *attention required*, *physical effort*, *time pressure*, *distraction* or *frustration* associated with trying to do the task while continuing to drive safely.

Radio - **Verbal** Interface to Select a Preset Station

• •
Low 0 1 2 3 4 5 6 7 8 9 10 High

Radio - Verbal Interface to Change from AM to FM and tune to a Specific Station

• •
Low 0 1 2 3 4 5 6 7 8 9 10 High

Navigation System - Entering a Destination Street Address

• •
Low 0 1 2 3 4 5 6 7 8 9 10 High

Navigation System – Canceling a Route Request

• •
Low 0 1 2 3 4 5 6 7 8 9 10 High

Blank-Back Number Task (listening to numbers only)

• •
Low 0 1 2 3 4 5 6 7 8 9 10 High

0-Back Number Task

• •
Low 0 1 2 3 4 5 6 7 8 9 10 High

1-Back Number Task

• •
Low 0 1 2 3 4 5 6 7 8 9 10 High

2-Back Number Task

• •
Low 0 1 2 3 4 5 6 7 8 9 10 High

Radio - **<u>Manual</u>** (button) Interface to Select a Preset Station

	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
Low	0		1		2		3		4		5		6		7		8		9		10	High

Radio - Manual Interface to Change from AM to FM and tune to a Specific Station

	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
Low	0		1		2		3		4		5		6		7		8		9		10	High

(**Note:** The text above was presented on a single "legal size" sheet of paper on a clipboard so that the participant was able to see all scales at the same time when making ratings.)

Appendix B: Radio Task Changes from Study 1 to 2

The radio task used in Study 1 was closely modeled on the "radio easy" and "radio hard" tasks employed in the CAMP Driver Workload Metrics Project (Angell et al., 2006). After the design of Study 1 was already in place, NHTSA released their proposed visual-manual distraction guidelines (NHTSA, 2012) which included adoption of a form of the radio-manual tuning reference task that differs slightly from the CAMP "hard" task in which the driver turns the radio on, switches the band type (AM/FM1/FM2), and then manually rotates the tuning knob to a specified frequency. In the NHTSA version, the radio is already assumed to be on (which is more in line with how The Alliance guidelines define the manual radio reference task).

Given the relevance of the NHTSA guidelines to future work in this area, the manual "hard" radio task was modified for Study 2 so that it conforms to the NHTSA guidelines (i.e. the radio is in the "on" state prior to the participant being asked to change the band setting and tune to a specific radio frequency). The NHTSA guidelines further specify that if the radio controls are part of an integrated vehicle display, the integrated display should be set so that the radio controls are not active; a test participant performs the action(s) necessary to make the radio controls active. A set-up step was thus added to the protocol for the manual radio tuning assessment to conform to this requirement as well.

While not a substantive change, the "hard" manual tuning task was changed from switching between FM1 and FM2 to switching between AM and FM, to be overtly in line with the NHTSA task description. Finally, some modifications were made in the specific station frequencies selected to support these changes.

A detailed listing of the task command structure employed in both studies is provided on the following pages. Some modifications in instructions were also made in consideration of the fact that Study 2 included a comparison of participants who were self-guided vs. participants who were given detailed training in how to complete each task. Speciffically, set-up instructions that provided specific guidance on how to complete certain steps were modified so that all participants had to depend on their memory of what they were instructed / discovered on their own during self-guided training to complete the tasks.

<u>"Hard" / Reference Manual Radio Task</u>

CAMP (radio off)

HARD 1: Your task is to **turn on** the radio, switch to the **FM** band, and tune to 107.5 FM.

HARD 2: Your task is to **turn on** the radio, switch to the **FM** band, and tune to 93.1 FM.

The tasks above require 3 manual steps: 1) pressing the power / volume button to turn the radio on, 2) pressing the FM band button, and 3) rotating the tuning knob to locate the specified station.

MIT Study 1 (radio off)

HARD 1: Your task is to **turn on** the radio, switch to **FM2**, and tune to 100.7

HARD 2: Your task is to **turn on** the radio, switch to **FM1**, and tune to 95.3.

The tasks above require 3 manual steps: 1) pressing the power / volume button to turn the radio on, 2) pressing the FM1 or FM2 band button, and 3) rotating the tuning knob to locate the specified station.

NHTSA (radio on)

Details of NHTSA's vision of the manual radio tuning reference task were provided in the draft guidelines document (NHTSA, 2012) (DS-BM.7 Reference Task; p. 158 in the manuscript version issued prior to publication in the Federal Register – this description was not carried over into the 2013 final guidelines document). Initially the radio is "On". If the radio controls are part of an integrated vehicle display, the integrated display should be set so that the radio controls are not active; a test participant performs the action(s) necessary to make the radio controls active, then changes the radio band selection from AM to FM (or vice versa). The participant then uses the radio tuning control to tune to the specified station with is approximately one-third of the band away from the start state. In the test vehicle studied, the tasks above require 3 manual steps: 1) pressing the [RADIO] button to make the AM and FM band buttons visible, 2) pressing the AM or an FM band button, and 3) rotating the tuning knob to locate the specified station.

MIT Study 2: (radio on)

HARD 1: Your task is to switch to **AM**, and tune to 1470.

HARD 2: Your task is to switch to **FM**, and tune to 100.7.

The tasks above require 3 manual steps: 1) pressing the [RADIO] button to make the AM and FM band buttons visible, 2) pressing the AM or an FM band button, and 3) rotating the tuning knob to locate the specified station.

Radio Pre-set Stations Settings Used in Study 2

Pre-set	FM1	FM2	AM
1	92.9	92.9	1030
2	94.5	94.5	1030
3	100.7	100.7	1030
4	104.1	104.1	1030
5	106.7	106.7	1030
6	107.9	107.9	1030

Preparatory Introduction to Manual Radio Tuning Evaluation On-Road

Study 1: "A task period is about to start. The tasks will involve operating the radio using the manual controls. You will be asked to do a number of tasks that you have already practiced. These include turning the radio on and off, changing stations using the preset buttons, switching between the AM and FM frequency bands, and manually locating a station using the tuning knob. There will be 10 to 60 second pauses between tasks. Do not begin a task until you hear the word 'begin'. When you have successfully completed a task, please immediately say the word 'done'. (Pause 3 seconds.) Your first task is to turn the radio on by pressing the volume knob. (Pause 2 seconds) Begin."

Study 2: NO CHANGES

Preparatory Introduction to Voice Radio Tuning Evaluation On-Road

Study 1: "A task period is about to start. The tasks will involve operating the radio using the voice interface. You will be asked to do a number of tasks that you have already practiced. These include turning the radio on, changing stations using the preset commands, switching between the AM and FM frequency bands, and requesting specific stations by frequency number. To turn the radio off, press the volume knob; for all other interactions, use the voice interface. There will be 10 to 60 second pauses between tasks. Do not begin a task until you hear the word 'begin'."(Pause 3 seconds.)<u>Your first task is to turn the radio on by pressing the push-to-talk button on the steering wheel and saying "radio-on</u>". (Pause 2 seconds) Begin."

Study 2: "A task period is about to start. The tasks will involve operating the radio using the voice interface. You will be asked to do a number of tasks that you have already practiced. These include turning the radio on, changing stations using the preset commands, switching between the AM and FM frequency bands, and requesting specific stations by frequency number. To turn the radio off, press the volume knob; for all other interactions, use the voice interface. There will be 10 to 60 second pauses between tasks. Do not begin a task until you hear the word 'begin'."(Pause 3 seconds.)<u>Your first task is to turn the radio on using voice command</u>. (Pause 2 seconds) Begin."

Manual Radio Tasks in Study 1

Recorded Audio Command	*Task Level*	*Action*
1. Your first task is to turn the radio on by pressing the volume knob. Begin.	(set-up task)	
2. Your task is to change the radio to preset-1. Begin.	Easy	goes to 92.9
3. Your task is to change the radio to preset-5. Begin.	Easy	goes to 106.7
4. Please press the Radio button, change to AM and then turn the radio off. Begin.	(set-up task)	
5. Your task is to turn on the radio, switch to FM2, and tune to 100.7. Begin.	Hard	tune to 100.7
6. Your task is to turn off the radio. Begin.	(set-up task)	
7. Your task is to turn on the radio, switch to FM1, and tune to 95.3. Begin.	Hard	tune to 95.3

At completion, participant is prompted to turn off the radio and then informed that the task is complete and to continue driving.

Manual Radio Tasks in Study 2

Recorded Audio Command	*Task Level*	*Action*
1. Your first task is to turn the radio on by pressing the volume knob. Begin.	(set-up task)	
2. Your task is to change the radio to preset-1. Begin.	Easy	goes to 92.9
3. Your task is to change the radio to preset-6. Begin.	Easy	goes to 107.9
4. Please press the volume knob to turn the radio off and then press it again to turn it back on. Begin.	(set-up task)	
5. Your task is to switch to AM, and tune to 1470. Begin.	Hard	tune to 1470
6. Please press the volume knob to turn the radio off and then press it again to turn it back on. Begin.	(set-up task)	
7. Your task is to switch to FM, and tune to 100.7. Begin.	Hard	tune to 100.7

At completion, participant is prompted to press [RADIO], [AM], [preset-4], and [FM]; this resets the default AM station to 1030 and then changes the default to FM. Finally, they are instructed to turn off the radio and then informed that the task is complete and to continue driving.

- The most overt change in Study 2 was the change to have the radio on at the start of the "hard" manual radio tuning tasks. This was done to be consistent with NHTSA guidelines.
- The step of turning the radio off and then on is done to reset the center console display so that the participant always needs to press the [RADIO] button to make the AM &FM soft buttons visible. This conforms to the NHTSA guidance the states that if the radio controls are part of an integrated vehicle display, the integrated display should be set so that the radio controls are not active; a test participant performs the action(s) necessary to make the radio controls active (p. 158).

Voice-Radio Tasks in Study 1

	Recorded Audio Command	*Task Level*	*Action*
1.	Your first task is to turn the radio on by pressing the push-to-talk button on the steering wheel and saying "radio-on". Begin.	(set-up task)	
2.	Your task is to change the radio to preset-1. Begin.	Easy	goes to 92.9
3.	Your task is to change the radio to preset-5. Begin.	Easy	goes to 106.7
4.	Please press the push-to-talk button, say Radio-AM and then turn the radio off. Begin.	(set-up task)	goes to AM 1030
5.	Your task is to turn the radio on using the push-to-talk button and request FM 100.7. Begin.	Hard	tune to 100.7
6.	Your task is to turn off the radio. Begin.	(set-up task)	
7.	Your task is to turn the radio on using the push-to-talk button and request FM 95.3. Begin.	Hard	tune to 95.3

At completion, participant is prompted to turn off the radio and then informed that the task is complete and to continue driving.

Voice-Radio Tasks in Study 2

	Recorded Audio Command	*Task Level*	*Action*
1.	Your first task is to turn the radio on using voice command. Begin.	(set-up task)	
2.	Your task is to change the radio to preset-1. Begin.	Easy	goes to 92.9
3.	Your task is to change the radio to preset-6. Begin.	Easy	goes to 107.9
4.	Your task is to request AM 1470. Begin.	Hard	tune to 1470
5.	Your task is to request FM 100.7. Begin.	Hard	tune to 100.7

At completion, participant is prompted to press the voice button and say the command "1030" followed by "yes"; this resets the default AM station to 1030. They are then instructed press the voice button and say the command "FM" and "yes" to return the default to FM mode. Finally, they are instructed to turn off the radio and then informed that the task is complete and to continue driving.

- The first prompt was modified to remove specific instruction on how to turn the radio on using voice command since this might serve as a potential aid to the self-trained group. This had the potential to make this initial step slightly more challenging for the trained group compared to Study 1; however, this was not a task that was being formally assessed and this change was anticipated to be largely inconsequential. In fact, self-reported workload for both the "easy" and "hard" voice tasks was found to be nominally lower in Study 2.
- The most overt change was the removal of the set-up steps before each of the "hard" task trials. These were no longer needed since the hard task no longer included turning on the radio.
- At the completion of the formal assessment tasks, the participant was guided through steps necessary to place the radio in a specified configuration to support any subsequent radio task trials.

Appendix C: Navigation Task Changes from Study 1 to 2

As discussed in the main body of this report, the prompt wording for the navigation tasks was modified in Study 2 to remove explicit instruction on how to complete the tasks. This was done to since one of the goals of the study was to evaluate to whether participants who were given structured training in how to complete the tasks varied in their behavior from those who were solely dependent on their own exploration of the DVI to discover how to complete the specified tasks. The text of the recorded preparatory instructions and specific task prompts used in the two studies is provided on the following page. Underlying is used to highlight differences between the studies.

Navigation Tasks MIT Study 1

Preparatory Introduction: "A task period is about to start. The tasks will involve operating the navigation system using the voice interface. You will be asked to do two kinds of tasks that you have already practiced. These consist of using the command, 'destination street address' to begin an address entry and then later cancelling the route. There will be pauses of approximately one minute between tasks. Do not begin a task until you hear the word 'begin'. There will be a one minute pause before the first task is presented."

Recorded Audio Command	*Task Level*
1. Your task is to enter the destination address: 177 Massachusetts Avenue, Cambridge, Massachusetts. Begin.	Hard
2. Your task is to cancel the route using the command 'Navigation Cancel Route'. Begin.	Easy
3. Your task is to enter the destination address: 293 Beacon Street, Boston, Massachusetts. Begin.	Hard
4. Your task is to cancel the route using the command 'Navigation Cancel Route'. Begin.	Easy
5. The task is complete. Please continue driving.	-

Navigation Tasks MIT Study 2

Preparatory Introduction: "A task period is about to start. The tasks will involve operating the navigation system using the voice interface. You will be asked to do two kinds of tasks that you have already practiced. These consist of entering a destination using a stress address and then later cancelling the route. There will be pauses of approximately one minute between tasks. Do not begin a task until you hear the word 'begin'. There will be a one minute pause before the first task is presented."

Recorded Audio Command	*Task Level*
1. Your task is to enter the destination address: 177 Massachusetts Avenue, Cambridge, Massachusetts. Begin.	Hard
2. Your task is to cancel the route you entered. Begin.	Easy
3. Your task is to enter the destination address: 293 Beacon Street, Boston, Massachusetts. Begin.	Hard
4. Your task is to cancel the route you entered. Begin.	Easy
5. The task is complete. Please continue driving.	-

- As highlighted above, the preparatory introduction for Study 2 was modified to remove the explicit reminder that the command 'destination street address' is used as a short-cut to enter an address into the navigation system. This was done so that the 'self-guided training' group would continue solely dependent upon their own self-exploration of the system to determine how to complete this task. Similarly, in Study 1, participants were explicitly prompted on the command to use to cancel a route so that we could assess the ease of doing the task as opposed to testing how well the participant remembered the command. This was modified in Study 2 to again avoid explicitly revelling to the 'no training' group the explicit command to use if they had not discovered it on their own.

Appendix D: Baseline Reference Period

As was discussed in the report for Study 1 (Appendix E), the use of a single reference period immediately prior to a defined secondary task period is a common method by which dynamic measures such as physiology, glance behavior, and driving performance metrics can be considered relative to "single task" driving. One of the challenges for real-world studies is the reality that external events, such as truck passing close by to the experimental vehicle, or internal events, such as thoughts about personal conflicts, will influence the arousal level and driving performance of some percentage of participants during any given reference point. This can have a measurable impact on the estimated change in behavior associated with single task vs. dual task conditions as we have discussed previously (Reimer, Mehler, Wang, et al., 2012).

In the present study, we collected a series of 5 baseline periods, one prior to each task block (navigation, manual radio, voice radio, standard n-backs, and an exploratory self-paced n-back task following the primary tasks) and combined these for purposes of establishing a broad single task driving reference. Each individual baseline period was 2 minutes in duration and the combined baseline value for each variable of interest represents the mean across the 5 baselines, i.e. representing a cross-section of 10 minutes of single task driving.

Structural Relationship between Baseline Periods and Task Blocks:

Task Block 1		
Baseline (single task driving)	120	seconds
Separation	30	
Introductory Instructions for Task Block 1	~	
Tasks x – y	~	
Separation	30	
Task Block 2		
Baseline (single task driving)	120	
Separation	30	
Introductory Instructions for Task Block 2	~	
Tasks x – y	~	