SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Impact of Smart Phones' Interaction Modality on Driving Performance for Conventional and Autonomous Vehicles

Francesca M. Favarò, PhD

MINETA TRANSPORTATION INSTITUTE
LEAD UNIVERSITY OF
Mineta Consortium for Transportation Mobility

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the four-university Mineta Consortium for Transportation Mobility, a Tier 1 University Transportation Center funded by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology (OST-R), the California Department of Transportation (Caltrans), and by private grants and donations.

MTI's transportation policy work is centered on three primary responsibilities:

Research

MTI works to provide policy-oriented research for all levels of government and the private sector to foster the development of optimum surface transportation systems. Research areas include: bicycle and pedestrian issues; financing public and private sector transportation improvements; intermodal connectivity and integration; safety and security of transportation systems; sustainability of transportation systems; transportation / land use / environment; and transportation planning and policy development. Certified Research Associates conduct the research. Certification requires an advanced degree, generally a Ph.D., a record of academic publications, and professional references. Research projects culminate in a peer-reviewed publication, available on TransWeb, the MTI website (http://transweb.sjsu.edu).

Education

The Institute supports education programs for students seeking a career in the development and operation of surface transportation systems. MTI, through San José State University, offers an AACSB-accredited Master of Science in Transportation Management and graduate certificates in Transportation Management, Transportation Security, and High-Speed Rail Management that serve to prepare the nation's transportation managers for the 21st century. With the active assistance of the California Department of Transportation (Caltrans), MTI delivers its classes over a state-of-the-art videoconference network throughout the state of California and via webcasting beyond, allowing working transportation professionals to pursue an advanced degree regardless of their location. To meet the needs of employers seeking a diverse workforce, MTI's education program promotes enrollment to under-represented groups.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and journals and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation- related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

REPORT 20-04

IMPACT OF SMART PHONES' INTERACTION MODALITY ON DRIVING PERFORMANCE FOR CONVENTIONAL AND AUTONOMOUS VEHICLES

Francesca M. Favarò, PhD

Student Co-Authors:
Marcel Derichs
Sky Eurich
Syeda Rizvi
Sara Akthar
Shivangi Agarwal
Kristina Davtyan
Nazanin Nader
Kiarash Behzad
Sumaid Mahmood

January 2020

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 20-04	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Impact of Smart Phones' Interaction Modality on Driving Performance for Conventional and Autonomous Vehicles		**5. Report Date** January 2020
		6. Performing Organization Code
7. Authors Francesca M. Favarò, Ph.D., https://orcid.org/0000-0003-0239-0770		**8. Performing Organization Report** CA-MTI-1813
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		**10. Work Unit No.**
		11. Contract or Grant No. 69A3551747127
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology University Transportation Centers Program 1200 New Jersey Avenue, SE Washington, DC 20590		**13. Type of Report and Period Covered** Final Report
		14. Sponsoring Agency Code
15. Supplemental Notes DOI: 10.31979/mti.2020.1813		

16. Abstract

Distracted driving related to cell phone usage ranks among the top three causes of fatal crashes on the road. Although forty-eight of 50 U.S. states allow the use of personal devices if operated hands-free and secured in the vehicle, scientific studies have yet to quantify the safety improvement presumed to be introduced by voice-to-text interactions. Thus, this study investigated how different modes of interaction of drivers with a smart phone (i.e., manual texting vs. vocal input) affect drivers' distraction and performance in both conventional and semi-autonomous vehicles. The study was executed in a full-car integrated simulator and tested a population of 32 drivers. The study considered two scenarios: (1) conventional manual driving in a suburban environment with intersection stops; and (2) control takeover from an engaged autonomous vehicle that reverted to manual driving at a highway exit. The quality of execution of maneuvers as well as timing and tracking of eye-gaze focus areas were assessed in both scenarios. Results demonstrated that while participants perceived an increased level of safety while using the hands-free interface, response times and drift did not significantly differ from those manually texting. Furthermore, even though participants perceived a greater effort in accomplishing the text reply through the manual interface, none of the measured quantities for driving performance or eye-gaze focus revealed statistical difference between the two interfaces, ultimately calling into question the assumption of greater safety implicit in the laws allowing hands-free devices.

17. Key Words Driver performance, distraction, text messaging, autonomous vehicle handover, reaction time	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 71	**22. Price**

Form DOT F 1700.7 (8-72)

012920

ACKNOWLEDGMENTS

The contents of this report reflect the views of the authors, who are responsible for the facts and accuracy of the information presented herein. This report does not necessarily reflect the official views or policies of the U.S. government, the State of California, or the Mineta Transportation Institute, who assume no liabil-ity for the contents or use thereof. This report does not constitute a standard specification, design standard, or regulation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The National Highway Traffic Safety Administration estimates that 94% of U.S. crashes are imputable to human error, with distraction related to cell phone usage in the top three causes of incidents on the road. Forty-eight out of 50 U.S. states allow the use of personal devices if operated hands-free and secured in the vehicle. This seemingly convenient solution carries with it the implicit assumption that vocal interactions with a personal device are less distracting and do not disrupt the driver's attention from the road as much. Yet scientific studies to quantify the safety improvement brought along by voice-to-text interactions are lacking within the literature. This study hence aimed at investigating how different modes of interaction of drivers with a smart phone (i.e., manual texting vs. vocal input) affect a driver's distraction and performance in both conventional and semi-autonomous vehicles.

The study was executed in a full-car integrated simulator, and it tested a population of 32 drivers. Two scenarios were considered:

1. A suburban scenario with a regulated intersection stop (i.e., traffic light), subject to the reception of two text messages, and driven in conventional manual driving;

2. A highway scenario with exit lane merging, subject to the reception of one text message, and driven first autonomously by the vehicle and then manually by the driver through a control takeover maneuver following the autonomy disengagement when the vehicle was approaching the highway exit.

Each scenario was repeated twice, once asking participants to reply through manual texting, and once through vocal dictation. Different variables were measured for the two scenarios, which can be summarized in the following categories:

- Time-to-Input variables: these were time-related metrics that measured drivers' response times to particular stimuli. Those included time-to-throttle after the traffic light turned green in the suburban scenario and time-to-steering/throttle/braking after the autonomy disengagement for the highway scenario;

- Vehicle trajectory variables: these were vehicle-related metrics related to distance from a vehicle in the front for the suburban scenario (stop-and-go traffic at the lighted intersection), and lateral offset and drift behavior for the highway scenario;

- Gaze dwelling and focus: these were eye-tracking related variables aimed at quantifying the time spent looking at the phone versus at the road;

- Texting behavior: these were time-related metrics that quantified how quickly a driver engaged with the cell phone after text reception, how long it took to compose the message, and the length of the overall reply. Speed of the vehicle during cell phone engagement was also measured;

- Subjective measures: these were self-assessed metrics that each participant

provided for each test through pre-test, post-test, and in-between scenario surveys. Surveys queried participants on the levels of perceived safety, effort, mental workload, trust, comfort, and human-machine interfaces.

All variables were measured in the situation where driving was interleaved with the secondary non-driving task of receiving a text that required a reply, once through the manual texting interface, and once through the vocal reply interface. The manual versus vocal interface was the only independent variable tested in this study. Each participant experienced four tests (2 scenarios × 2 interfaces), which were randomized in order to reduce the effect of learning.

Two-tailed t-tests were used to analyze statistical difference for all the dependent variables investigated as a function of the two tested interfaces (i.e., manual text reply and vocal reply). The main conclusion of the study rests with the different nature of those variables that were found to be statistically significant and those that were not. In particular, two types of variables were found to be statistically significant:

1. **Texting-related variables:** composition time for a text was found to be lower for the vocal interface. Length of text in terms of characters contained was similar for both interfaces, but the vocal interface (as intuitively thought) gave a shorter time of text composition. In other words, people were faster at composing a reply through the vocal interface. Furthermore, for the suburban scenario, marginal significance was found for the time of first-click, or time-to-interaction, with people more likely to tap on the phone sooner with the manual interface.

2. **Subjective measures:** lower perceived effort was found for the vocal interface, and participants expressed a clear preference for the vocal interface, indicating that the level of compromised safety was higher for the manual interface. In other words, participants felt safer and less tired by the vocal composition interface.

Despite the seeming advantages outlined above (i.e., shorter texting times and higher perceived safety with lower effort), the vocal interface did not provide any statically significant improvement seen in participants' performance with respect to the objective engineering metrics investigated in the study, including response times and drift/lateral offset. In other words, while participants preferred the vocal interface and pointed to the fact that it was a safer option, their actual performance in the vehicle did not statistically differ between the manual and the vocal interface. This conclusion has implications for the current state of driving regulations in the U.S., which overwhelmingly enforce a texting ban while making no stipulations regarding the use of voice-reply technology. The authors thus bring forward the careful recommendation to further study hands-free interfaces and to be wary of the technological promises of vocal engagement interfaces coming out for infotainment purposes. Nevertheless, the results of this report are to be considered preliminary and limited in scope, given the small sample tested and the non-generalizable nature of the specific scenario and human-machine interfaces considered here.

I. INTRODUCTION

The National Highway Traffic Safety Administration (NHTSA) estimates that 94% of U.S. crashes are imputable to human error, with distraction related to cell phone usage in the top three causes of incidents on the road.[1] A recent report by Zendrive for NHTSA estimates that, regardless of the regulations in place, over 60% of the driver's population engage in cell phone-based activities while driving at least once per day.[2] Currently there is no national ban on texting or using a wireless phone while driving, but the majority of U.S. states have passed laws banning texting or wireless phones or requiring hands-free use of wireless phones while driving.[3] Figure 1 provides a visual summary of current U.S. requirements on texting while driving regulations.[4] California is 1 out of 21 states that carries a specific state ban on hand-held devices for all drivers.

Cell Phone Use and Texting While Driving Laws

Note: Many states that have passed hand-held phone bans and texting bans provide for various exemptions for emergencies, law enforcement personnel, emergency medical technicians, firefighters, state DOT employees, etc.

Source: National Conference of State Legislatures, 2018.

- Primary enforcement of texting ban for all drivers
- Secondary enforcement for texting ban for all drivers
- Texting ban for novice/beginner drivers
- No testing ban in the state
- Hand-held phone use ban for all drivers

Figure 1.   Summary Map of Cell Phone Bans by State, from the National Conference of State Legislation[5]

The decision to allow hands-free use while at the wheel carries with it the implicit assumption that vocal interactions with a personal device are less distracting and do not disrupt the driver's attention from the road as much as fully manual interactions. Yet there is a consensus within the research community on the need to establish clear and quantitative connections between the usage of smart phones and driving performance.[6]

The use of portable devices has also been in the spotlight in relation to recent crashes of semi-autonomous vehicles,[7] where current regulations mandate the driver to monitor

the outside environment at all times.[8] With the advent of autonomous vehicles, it is more important than ever to assess drivers' capability to monitor the surrounding environment and to take over control when needed, while at the same time being engaged in activities other than driving. New technologies that leverage voice commands for interaction with portable devices are now becoming standard for vehicles deployed on the market: through Apple "carplay," a smart phone can be integrated within the vehicle, and voice commands allow the driver to respond to texts, provide desired navigation applications, and set music and radio stations.[9] While voice commands have been advertised by tech companies as a safety improvement effective for reducing distraction,[10] scientific evidence to quantify the safety improvement brought along by voice-to-text interactions is lacking within the literature. Moreover, novel technology is also available to quantify drivers' gaze focus areas and eye tracking and timing, allowing for more precise results than those traditionally obtained by direct observation of drivers' behavior.

This study thus aimed at investigating how different modes of driver interaction with a smart phone (i.e., manual texting versus vocal input) affect drivers' distraction and performance in both conventional and semi-autonomous vehicles. The study was executed in a full-car integrated simulator, and it tested a population of 32 drivers. Two scenarios were considered: 1) conventional manual driving in a suburban environment with intersection stops; and 2) control takeover from an engaged autonomous vehicle that reverted to manual driving at a highway exit. For both scenarios, quality of execution of maneuvers as well as timing and tracking of eye-gaze focus areas were assessed.

II. BACKGROUND AND LITERATURE

A substantial body of literature exists from 2000 to 2018 that addresses distracted driving and engagement with personal devices.[11] Previous research shows mixed results, showing manual texting to be clearly correlated to impairment of normal driving activities,[12] but also revealing the existence of circumstances under which the use of phones enhances alertness due to more expected threats.[13] Caird, Willness, Steel, and Scialfa[14] conducted a meta-analysis on the effects of cell phones on driver performance across multiple studies. They found that when phone-related secondary non-driving tasks are accounted for, a mean increase in reaction times to external stimuli presents, and these researchers quantified it at around 0.25 seconds for all types of phone-related secondary tasks. They also noted that such a performance decrease was probably underestimated in their study, with the true behavior of drivers in normal conditions (beyond those of the experimental setup) possibly worsened. Such "true behavior" can be assessed through studies that go under the name "naturalistic." One of those, by Fitch et al., also pointed to the strong correlation between manual texting and the impairment of normal driving activities.[15] Fitch conducted a naturalistic driving study in which 204 drivers were recorded continuously for one month. This is substantial, because it shows the effects of cell phone use in true-to-life driving scenarios rather than lab-based studies. Furthermore, Bakhit, Guo, and Ishak's study used naturalistic driving data (SHRP2 Naturalistic Driving Study) to assess the crash risk associated with different secondary tasks.[16] They found that cell phone texting and reading were some of the highest crash risk factors among all of the tasks they assessed.

Moreover, in conjunction with automation, research has shown that drivers of highly automated vehicles (i.e., vehicles capable of assisting with lane centering and steering) are more likely to engage in secondary activities while driving.[17] Current regulations require drivers of semi-autonomous vehicles to remain vigilant while automation is engaged, monitoring the driving environment at all times for better chances of successful recovery of control should the automation system request that the human driver regain manual control of the vehicle. It is thus more important than ever to obtain clear results regarding how the available technology can work in mitigating driver distraction.

In 2011, NHTSA conducted a national survey on drivers' distraction and assessed that at every point in time, over 660,000 drivers are talking over a hand-held phone. While a new national survey is in the making, the 2011 data showed dire statistics for driver distraction related to cell phone usage, with one in two drivers always answering incoming calls while at the wheel, one in four drivers placing calls out, and over 21,000 crashes occurring due to phone-related distraction.[18] Both NHTSA's survey of 2011 and Fitch's study of 2013 for the Department of Transportation focused on calls, with text messaging identified as a growing trend and estimated as affecting at least 14% of the driving population. Over the years, text messaging and emails have started to replace the traditional phone call, with new estimates for cellular texting ranging in number from an occurrence rate of 32% (active reply) to 41% (reading) regardless of texting bans.[19] Indeed, 21 out of 50 states in the U.S. have a ban on hand-held devices, and 48 out of 50 ban text messaging.[20]

Previous research has focused on comparing driver performance in the absence of non-driving secondary tasks with that achieved when programmed distraction is included into

the driving environment. The interested reader is referred to the encompassing literature review provided by Young and co-authors on drivers' distraction due to engagement with personal devices.[21] The present research will instead focus on *how* such interaction with a secondary non-driving task is carried out. To the authors' knowledge, the literature on comparisons of manual and speech-based interfaces is mostly inconclusive. Baròn and Green provided a review of the studies executed on the topic before 2006, and highlighted how it was not possible to determine (across a span of 15 studies) which interface was better, given the differences in implementations and the unknown nature of the interaction for different drivers.[22] In more recent years, and thus leveraging more up-to-date technology, only one other study was found to compare touch-based versus voice-based messaging interfaces for drivers' performance in selected maneuvers.[23] That study concluded that both manual-based and voice-based interfaces caused distraction, and it was inconclusive in showing whether one could be considered statistically better than the other, given that the two interfaces tested provided results that were hard to compare due to the different duration for task completion for the two interfaces. Furthermore, the study was executed within the limited task of manual conventional driving to follow a lead vehicle so that it presents significant limitations compared to the purpose of the present research.

The present work thus leverages three fundamental ingredients that were not encountered in the current literature and have not been combined in a unique study before: 1) a careful quantification of the improvement occurring when vocal commands are used instead of manual touch, thus comparing two different modes of interactions with smart phones for the same activity; 2) application for both conventional vehicles and autonomous vehicles in a quantification of quality and timing on control regain after failure of the autonomous technology in the presence of non-driving secondary tasks; and 3) usage of state-of-the-art eye tracking technology to assess gaze dwelling on cell phone and quantify gaze focus areas for the two interfaces.

III. METHODOLOGY

EXPERIMENTAL SETUP

This study employed human-in-the-loop (HITL) simulation, where participants sat in a fully-integrated car simulator. The static simulator, shown in Figure 2, featured a real BMW series 6 rented from FKA Silicon Valley.

Figure 2.   The HITL Simulator Owned by FKA Silicon Valley, A Subsidiary of the German company FKA GmbH

The simulator employed in this study is NHTSA-compliant for human-machine interface (HMI) evaluations and is capable of handling manual control by the driver as well as automated driving.

The simulation environment used a Linux-based simulation framework called Virtual Test Drive (VTD) by Vires Simulationstechnologie GmbH in version 2.1.0. Open standards (OpenDRIVE® and OpenSCENARIO) were used for road and scenario creation. The simulated driving environment was displayed on a 220-degree surround projection screen with a resolution of 1080 × 1920 and a refresh rate of 60 Hz. The simulator showed road and traffic information projected in front of the car and through a rear-view mirror behind the vehicle. A three-way split rear-projection wall provided the projection for side and rear-view mirrors.

The inside of the vehicle was equipped with a central console with a 10.2-inch screen, as well as an analogic dashboard on which the driver could read the speed of the vehicle. Drivers could adjust the seats' positions, seat belt height, side mirror headings, and rear-view mirrors according to their preferences. Furthermore, a 4.3-inch Nexus 6 touch-screen smart phone running Android OS Version 5 was used for the texting task and to navigate the participants through the simulated city and highway environments.

The eye/gaze tracker system used four cameras rented from FKA headquarters in Aachen, Germany. Figures 3 and 4 show the setup of the cameras for eye tracking within the dashboard and the back of the vehicle.

Figure 3.   Shot of the Front Dashboard of the Vehicle

Figure 4. Detail of the Back Camera

Three infrared cameras (model BASLER acA640) were mounted across the front dashboard and focused on correctly tracking eye gaze; a fourth one, seen in Figure 4, helped with pinpointing head orientation and was mounted to the back of the driver's seat (model Watec WAT-233). A calibration process before each test helped ensure that eye movement was appropriately followed by the cameras. The calibration was done for both the front and back cameras, and it consisted of adjusting the three front cameras so they would fit the participants' head profile. Participants were instructed to directly stare at each camera for a few seconds until the software created a 3D map of their face and pinpointed head rotation and pupils' orientation to relate the measured orientation to the pre-specified gaze location known for each camera. For the back cameras, participants were instructed to gaze at nine test points, which were tracked by the front cameras while the back camera calibrated head rotation for each point.

Figure 3 also showcases the digital speedometer that kept track of the vehicle speed, as well as the 10.2-inch central console screen at the front-center of the vehicle. The central console displayed the mode of operation of the vehicle. Three possible modes of operations were displayed through icons depicted in Figure 5:

- Autopilot Inactive: indicated to the left of Figure 5; used for conventional manual driving;

- Autopilot Active: indicated in the middle of Figure 5; used for automated driving;

- Prepare to Takeover: indicated to the left of Figure 5; used as transitional state from the active autopilot to the inactive autopilot, to alert and prime the driver to get ready to take over control (additional details on the disengagement structure are presented later).
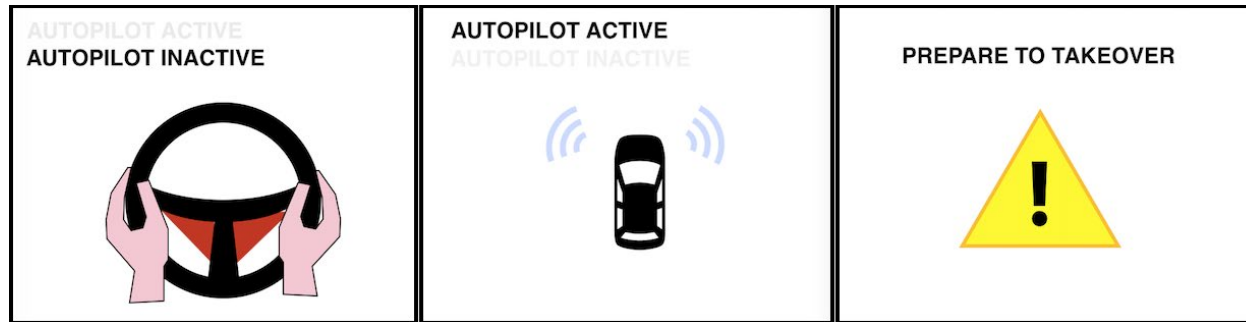
Figure 5.   Icons used to Display the Vehicle "Mode of Operation" Within the Central Console

Finally, Figure 3 also shows the cell phone holder employed in the study, which was mounted within the CD-reader slot beneath the air vents in the central-front area of the vehicle.

TEST STRUCTURE

Figure 6 shows how each test was structured schematically. The entire experience, from participant greeting to participant dismissal, took place over a duration of 1 hour and 15 minutes. A team of two researchers handled each test: one person was in charge of guiding the participant, sitting through the compilation of pre-test and post-test questionnaires, and sitting with the participant in the simulator vehicle during the test; the second researcher would sit in the control room and manage the simulation execution from the computer screens as well as monitoring the data logging process.
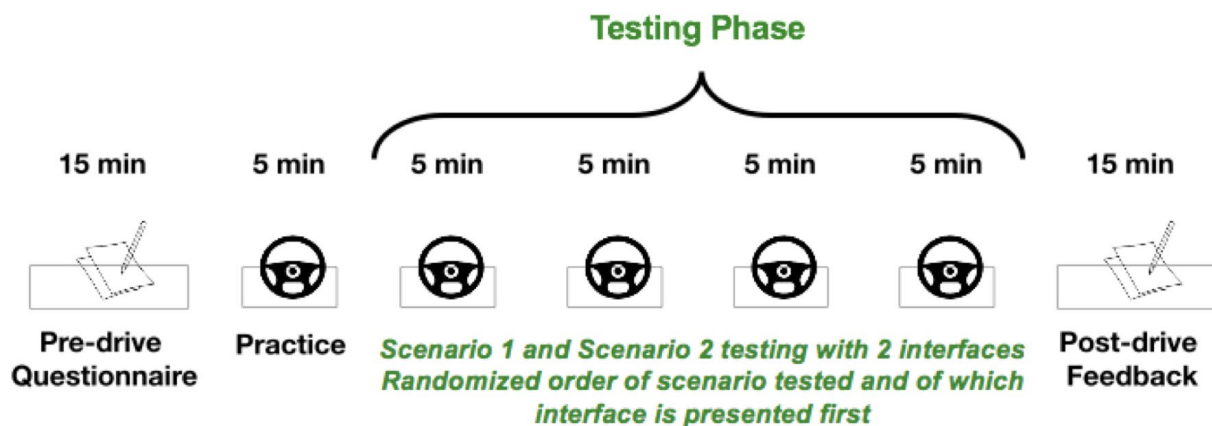
Figure 6.   Schematic Representation of the Test Structure

The test structure comprises seven components as follows:

1. Pre-drive questionnaire. Before starting the simulation, the researchers asked participants to fill out a pre-test questionnaire. The questionnaire included the necessary demographic information of participants as well as their driving history, authorized state of their driving license, their history of car accidents, the type of the car they drove, any autonomous features their car had, and, if it had autonomous features, the frequency with which they engaged them. The primary intention of the driving history section of this questionnaire was to gather information about participants' driving background and the ways in which it could be related to their views on autonomous driving. The questionnaire also assessed the participants' physical condition by asking them about the hours of sleep received the previous night and about any physical strain due to work activity; this was done in order to reject anyone potentially experiencing severe fatigue, which would affect the study results. Finally, the questionnaire asked participants about their overall attitude towards the test (excitement, nervousness, as well as trust in the technology).

2. Practice. Participants were given the opportunity to practice in the simulator to familiarize themselves with the vehicle employed in the study. This practice phase was executed in a highway simulated environment and lasted approximately 5 minutes. Each participant was offered to continue this phase until they were comfortable with the vehicle; all participants expressed comfort and none of them requested an extension of the practice phase. During the practice phase, the researcher sitting with the participant asked them to execute specific maneuvers in order to establish a simulated-driving baseline for each participant. The maneuvers were the same for all participants, and they took place in the same order. Specifically, participants were asked to change lanes, to accelerate and overtake another vehicle, to decelerate and change lanes, to keep an average speed of 60 mph, and then to decelerate to 55 mph and change lane to occupy the rightmost lane. Once those tasks were executed correctly, the participants were asked whether they were comfortable or not. When comfortable, the participants were then asked to take the next highway exit, a maneuver executed to establish a baseline in conventional manual driving to be used as a comparison for the takeover request the drivers would undergo during the highway test. For the entire practice phase, participants drove manually, without assistance from the autonomous technology.

3. Four tests. This experiment tested two environments (suburban and highway), each with two interfaces (manual texting and voice-to-text). This setup gives a combination of four tests, which were executed in random order to reduce the impact of learning effects. Each test had a duration of close to 5 minutes, with a brief interval between each of the four instances to complete a brief in-between-tests survey that assessed the perceived workload and gauged situational awareness during the simulation. Details on those surveys are provided later in this report.

4. Post-drive feedback. A final survey investigated the participants' HMI preferences—in particular in relation to the transition from automated to manual control for the highway scenario—and asked for overall suggestions for improvement. The final

section of the questionnaire investigated a number of human factors concerning participants' emotional and physical states, including any changes in trust in the technology, levels of comfort, levels of anxiety and perceptions of danger, and any nausea and motion sickness.

Participants were informed before the beginning of the test that they could withdraw from participation at any time, and they were constantly monitored for signs of unease and discomfort.

SCENARIO INFORMATION

Two environments were selected for testing within this project:

1. Suburban environment with intersection stops and traffic light, executed in conventional manual driving;

2. Highway environment, executed in automated driving followed by a request for manual takeover and manual driving through a highway exit.

The two scenarios were picked to ensure that the interface effect on driving performance could be quantified in different conditions. Additionally, the automated driving portion was limited to the highway scenario, given the current technological limitations of this feature, which is not yet allowed within urban environments. Both scenarios included a combination of the driving task and non-driving secondary tasks. The secondary tasks included following GPS instructions and replying to texts "when comfortable" through the provided interface (i.e., manual texting or vocal reply). Both scenarios were repeated twice to account for both types of interface. Furthermore, in between each repetition, a short survey was filled out. Details on those tasks are presented next.

Driving Task

During manual city scenarios, participants had to complete a series of left and right turns, and to stop at stop signs and at a red light while they maintained a posted speed limit indicated in the navigation interface and distance from other vehicles. Figure 7 provides an overview of the city grid used for the test. The simulated driving scenario consisted of a 1-mile-long single-lane suburban street grid.
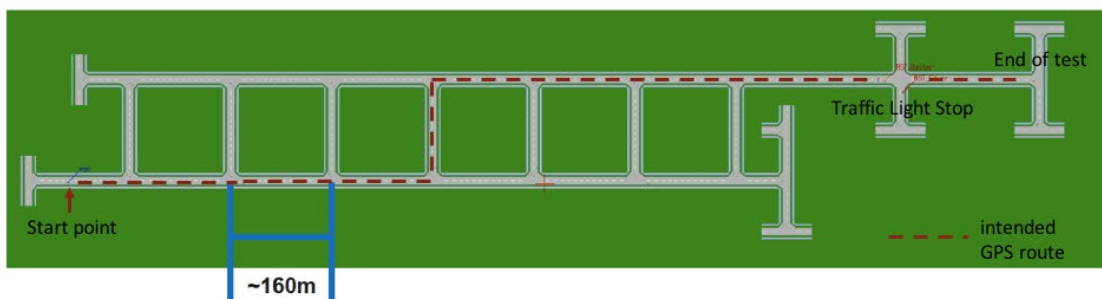
Figure 7.   Map of the Suburban Simulation Environment

The grid followed a precise geometry that would avoid drivers to nullify the test by non-compliance with the GPS instructions. The symmetric geometry was designed so that, should a driver take the wrong turn, the test did not need to be interrupted and restarted, but rather could continue after a "circle back around the block" to follow the intended path. None of the participants missed any turn.

During the fully-autonomous highway scenarios, participants were instructed to continue monitoring the outside environment while the vehicle was driven autonomously and to be ready to take over control of the vehicle if needed. When the GPS signaled the driver to take the next approaching exit, the driver was issued a warning within the central console display to prepare to resume manual control of the vehicle. This type of "priming" of the driver before the actual takeover happens is called a structured disengagement, meaning that the transition between the "autopilot active" and "autopilot inactive" is accounted for by an alert stage in which the vehicle is still driven autonomously, but the driver is made aware of the impending transition of control authority. Drivers were also pre-warned before the test that the autonomous technology could not handle highway exits, and that they would need to takeover control should the GPS indicate the need to take one. Also note that drivers had practiced the specific exit geometry during their practice phase. Traffic density was kept constant for all tests, for a total of 50 vehicles distributed within a 400-meter diameter from the test vehicle.

Figure 8 shows the map of the three-lane highway environment (approximate total length of 4 miles), with Figure 9 further providing a zoomed-in depiction of the geometry of the exit, along with two interest points of text reception and disengagement point (i.e., where the transition from automated to manual control happened).
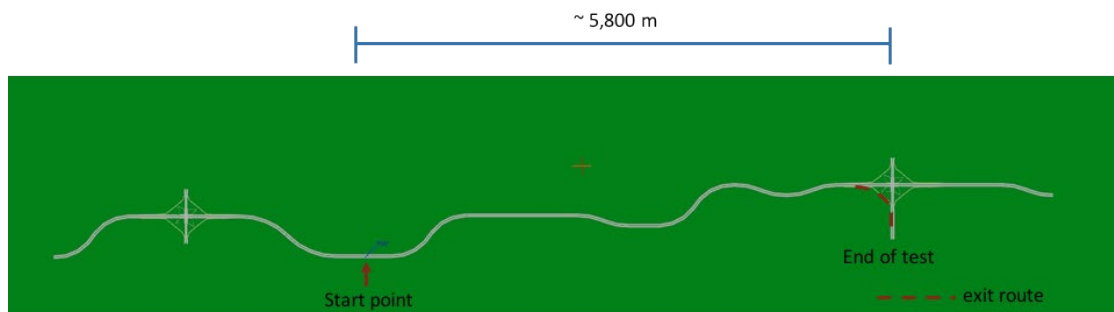
Figure 8.   Map of the Highway Simulation Environment
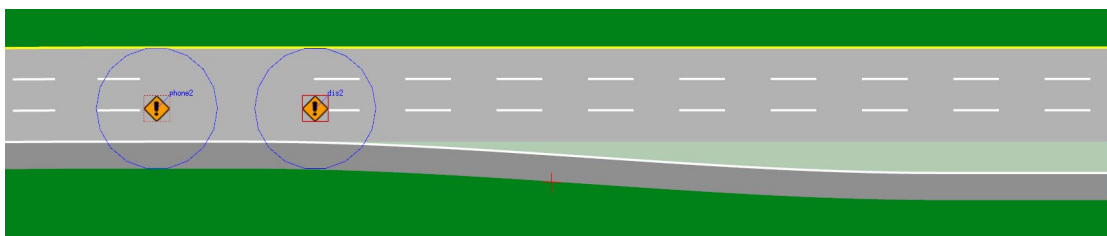
Figure 9.   Zoomed-In Map Section of the Highway Exit with Points of Interest

Secondary Tasks

Participants were instructed to follow the GPS instructions on the mounted Android phone during both the driving scenarios. A sample of how those instructions would look is provided in Figure 10. The appropriate navigation instruction was provided at the top of the screen, both spelled out and represented visually with arrows. The lane to occupy was also indicated, with a simplified view of the roads presented. Speed limit was included on the left side, and estimated time of arrival (ETA) in remaining minutes and seconds was provided at the bottom.
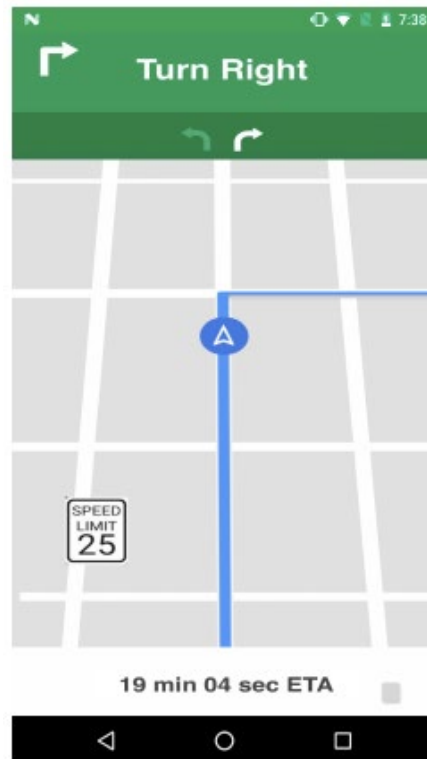
Figure 10. Sample GPS Instruction

A texting task was used to compare the distracting effects of manual versus vocal texting on driving performance. In the city scenario, two texts were introduced as a secondary non-driving task. The first text prodded the driver to provide an ETA (displayed on the GPS app), and it was introduced in conjunction with the traffic light turning red, slightly before approaching the intersection. While replying to the text, the participant would need to stop behind another vehicle, which was also stopped at the red traffic light. Upon the light turning green, the participant would then be required to resume driving, while a second text message would be received, asking whether the participant would prefer a cheese or pepperoni pizza for dinner. In the highway scenario, a single text was employed. The text, which again inquired about the ETA, was triggered right after the notification that the driver should "prepare to take over" for the vehicle movement. Fifteen seconds after the "priming" of the participant (i.e., after receiving the notification that the autopilot was about to disengage), the actual disengagement would be triggered. As shown in Figure 9, the driver would then be required to merge into the exit lane, while the received text awaited reply. Figure 11 and Figure 12 show how the interface looked for the manual reply and the vocal reply, respectively. Note that to reply vocally, the participant still had to first tap on the

microphone icon to begin recording the message. All participants were handed a training video one week before coming for the experiment, and they were asked for their level of comfort with the phone used in the experiment right before starting. All participants felt comfortable with the interface provided.
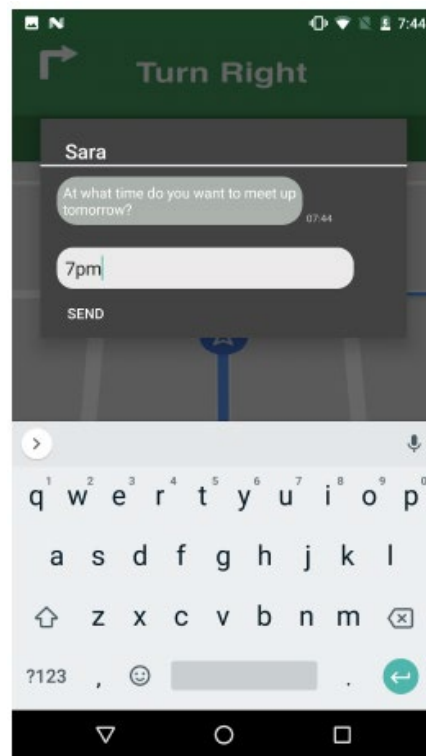
Figure 11. Sample of the Manual Texting Interface
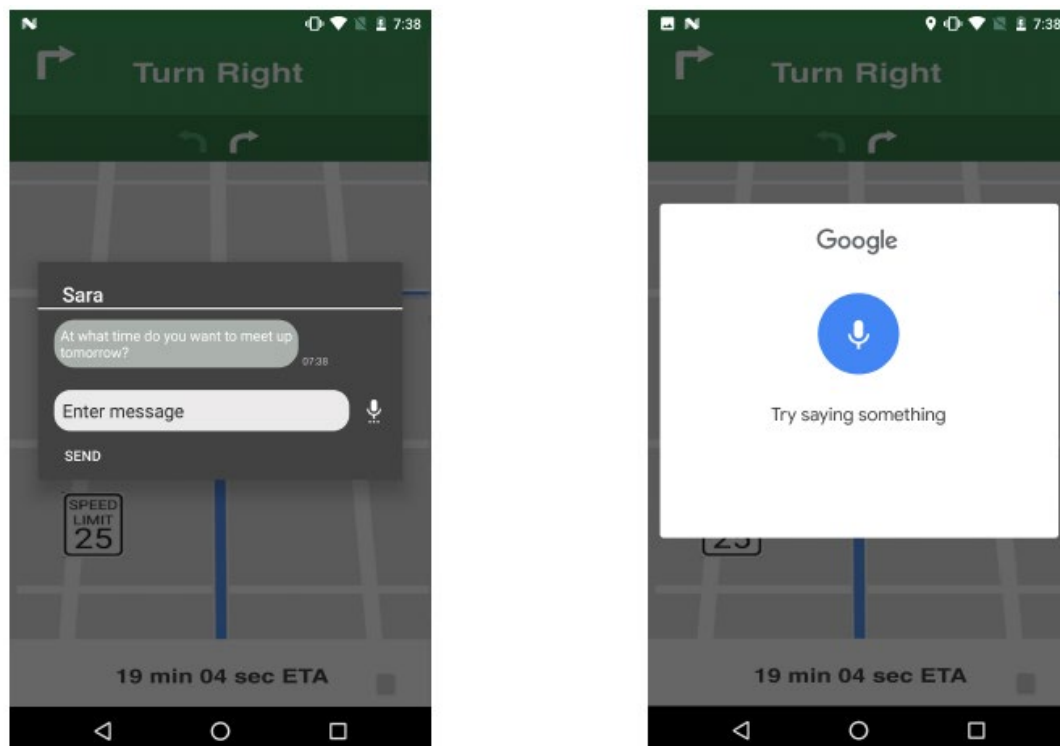
Figure 12. Sample of the Vocal Reply Interface

In-Between Survey

The in-between survey was composed of two parts: a task load index (TLX) analysis and a situational awareness portion. The TLX analysis served to compared participants' subjective ratings of the perceived mental difficulty of the driving tasks. The TLX analysis was developed by NASA in 1988,[24] and it consists of a paper-and-pen survey to be administered immediately after the end of a task (in the case of this experiment, at the end of each of the four driving scenarios), asking participants to rank the perceived "workload" associated to six subjective sub-scales. Workload is defined as "the perceived relationship between the amount of mental processing capability or resources and the amount required by the task."[25] The six sub-scales were rated on a 20-point range (from 0, minimum, to 20, maximum) for mental demand; physical demand; temporal demand; performance; effort; and frustration. Descriptions were provided for each item, and each participant completed a total of four TLX surveys: one for each driving scenario and texting interface combination. Furthermore, questions about recollection of scenario details (e.g., speed of travel, objects identified along the route) were included to test situational awareness. Appendix A provides a copy of the in-between survey for the interested reader.

DESIGN OF EXPERIMENT: VARIABLES

Table 1 provides a summary of the dependent variables measured for the study. Three categories of variables are identified for each scenario: those related to response times from the driver; those related to the vehicle trajectory; and those related to the gaze focus during the test.

Table 1.   Summary of Study Dependent Variables

Scenario	Time-to-Input Variables	Vehicle Trajectory Variables	Eye Tracking Variables
Suburban intersections and traffic light	Time-to-throttle: response time to the stimulus of the traffic light turning green in terms of acceleration input	Stopping distance behind a vehicle stopped at the red traffic light	▪ Time spent looking at the phone vs. at the road ▪ Maximum dwelling time of gaze on cell phone
Highway exit during automation disengagement	• Time-to-throttle • Time-to-steer • Time-to-brake (all computed from the disengagement point)	• Lateral offset from the center of the lane during takeover • Integral ratio to compare conventional driving to driving after disengagement	▪ Time spent looking at the phone vs. at the road ▪ Maximum dwelling time of gaze on cell phone

The only independent variable considered was the type of interface (i.e., manual or vocal). The scenario was not considered an independent variable itself, so that no multi-variable analysis of variance is carried out in this work. Two-tailed t-tests were used to assess whether there was a statistically significant difference among the variables recorded per Table 1 in the case of manual versus vocal interface.

PARTICIPANTS' SELECTION

After receiving approval from the Institutional Review Board (IRB) of San José State University (SJSU) in relation to human testing, participants were recruited via a flyer posted around the urban SJSU campus. Thirty-two participants completed the study

(16 men and 16 women, M = 26.8 years of age, SD = 6.5 years of age); eight additional participants did not complete the study due to motion sickness (6 participants), or excessive duration of the test (2 participants). Ten more participants had been recruited but cancelled at the last minute.

No incentive was provided for participation in the study. In order to be eligible for this study, all participants needed to have a valid U.S. driver's license and to have driven at least once in the 30 days prior to the test. All participants were screened for conditions, medical or otherwise, that would prevent the normal operation of a vehicle. Moreover, all participants were screened prior to participation to ensure normal or corrected-to-normal vision. This step had to occur because the calibration of the eye tracking cameras prevented the driver from using glasses (contact lenses were allowed). Heavy make-up and untrimmed facial hair were also prohibited in the study for camera calibration purposes, and participants with long hair were asked to tie it up.

Participant selection targeted the age range of 18 to 48. This range was chosen to mimic previous studies on the topic of cell phone usage, as evidence has shown that older populations are less likely to engage with personal devices.[26] Moreover, to achieve a better comparison between the two modes of interactions investigated in this work, the researchers targeted a young population of drivers who were more likely to be proficient and comfortable with both technologies (i.e., voice-based and touch-based). The population was equally split between male and female participants, to reflect the current equality witnessed in the U.S. driver population.[27]

Participants (anonymously) reported sending two text messages on average per drive (SD = 1.24). Figure 13 reports the breakdown and distribution of reported texting frequency per drive. Finally, participants were queried about current regulations in California for phone usage while at the wheel. A majority of the participants (65%) reported being aware of the hand-held ban for California, whereas 18% incorrectly reported the existence of an "all-ban" for all drivers. No participant reported that phone usage was unregulated and allowed without limitations.
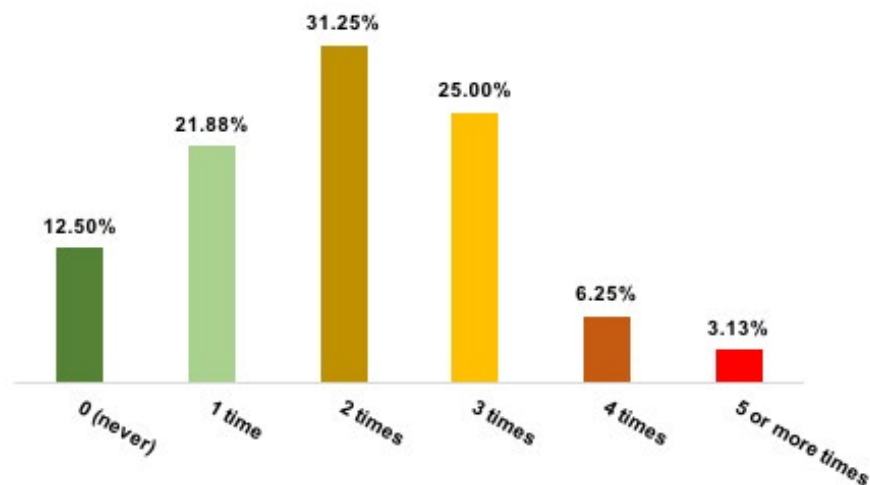
Figure 13. Breakdown of Reported Frequency of Text Messaging During Regular Drive

DATA COLLECTION

Simulator

The simulator central computer continuously logged the following quantities during all participants' activities (both practice phase and actual tests).

1. Road geometry

2. Test vehicle heading

3. Lateral lane offset (offset between the center of gravity of the vehicle and the center line of the lane of travel)

4. Speed of the vehicle

5. Steering angle

6. Brake pedal position (percentage between 0 and 100% of maximum vehicle braking capability)

7. Throttle input (percentage between 0 and 100% of maximum vehicle acceleration capability)

8. Test vehicle global position

9. Simulation time (elapsed from beginning of test) and frame number

10. Driving mode (automated, manual, or pre-warning phase)

11. Traffic light color (if traffic light is present)

12. Lane of travel

Road geometry is important for understanding the specific direction of travel, or "heading," that the road follows. The researchers measured the test vehicle heading, which is the heading of the vehicle the driver is in, in order to determine how divergent it was from the road geometry. Figure 14 illustrates the difference between these two outputs; the angle between these two headings is termed the "angular error" (the dashed grey line represents the center of the lane).
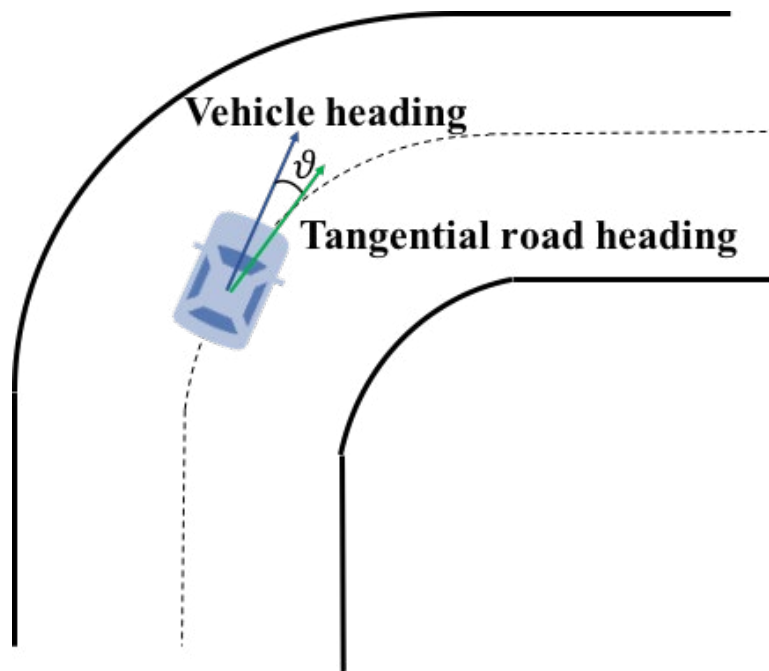
Figure 14. Representation of Angular Error

Lateral lane offset is the distance that the driver allowed the vehicle to drift from the center of their original lane during the disengagement, shown in Figure 15. Lateral offset results from accumulated angular errors that are not corrected by the driver.
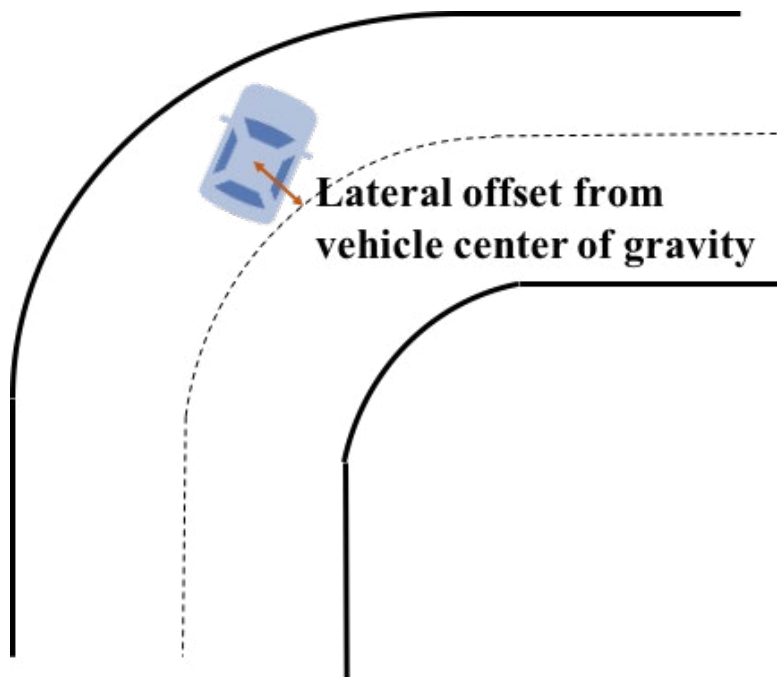
Figure 15. Representation of Lateral Offset

Steering angle was defined as the angle of the steering wheel, given as the angular difference from the neutral position. Brake and throttle outputs were given as percentages, between no depression (0%) and maximum depression (100%). The output was automatically generated as a .csv file.

Eye Tracker

The four cameras employed in the study recorded in real-time the upper body movement of the driver, and they were calibrated to provide a continuous stream of output through a classification algorithm, as explained next. Six categories of output were defined, according to Figure 16: (i) windshield; (ii) rearview mirror; (iii) cluster instrument (i.e., speedometer); (iv) center display; (v) handy display (i.e., the cell phone mounted underneath the central air vents); and (vi) radio.
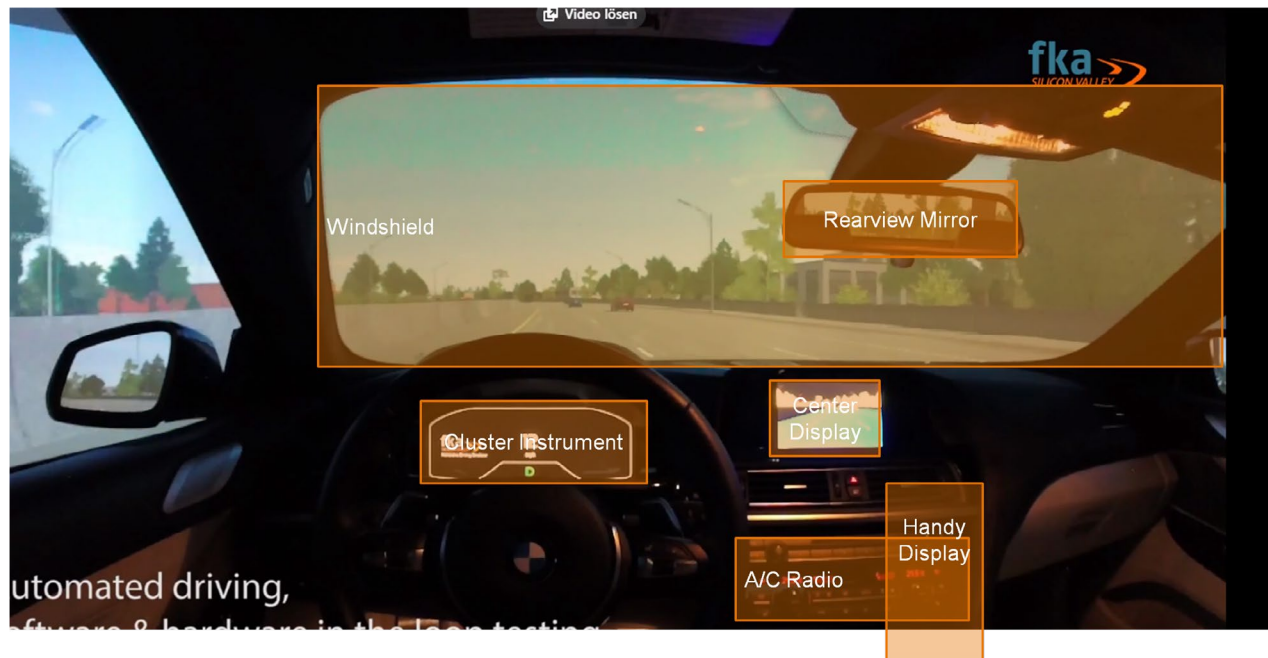
Figure 16. Eye Tracker Areas of Interest for Classification

At every point in time, the eye tracker recorded a timestamp associated to the recording, a time delay between the clock that ran with the cameras and that of the simulator, and a category output indicated as one of the six focus areas of Figure 16 or a seventh "unknown" category for when the participant was looking at regions that did not fall within those six categories (e.g., side-mirrors, driver's own lap). Also note that because of lighting conditions, the eye tracker was not always capable of recording an output. This situation would trigger an output labeled "invalid" if the accuracy threshold with which the eye tracker was assigning categories was not met. The software used for the classification process was Smart Eye Pro. This software defines a gaze quality index based on the blinks, saccades[28] and fixation values collected by the system. Any data with a gaze quality index lower than 0.25 is then considered invalid. Across the 32×4 = 128 successful tests, the valid values ranged anywhere from 49.3% to 97.4% with an average of 83.4%. Figure 17 shows the distribution of valid versus total collected data across all tests. The system was a state-of-the-art apparatus, and the validity results were in line with current literature on eye tracking.[29]
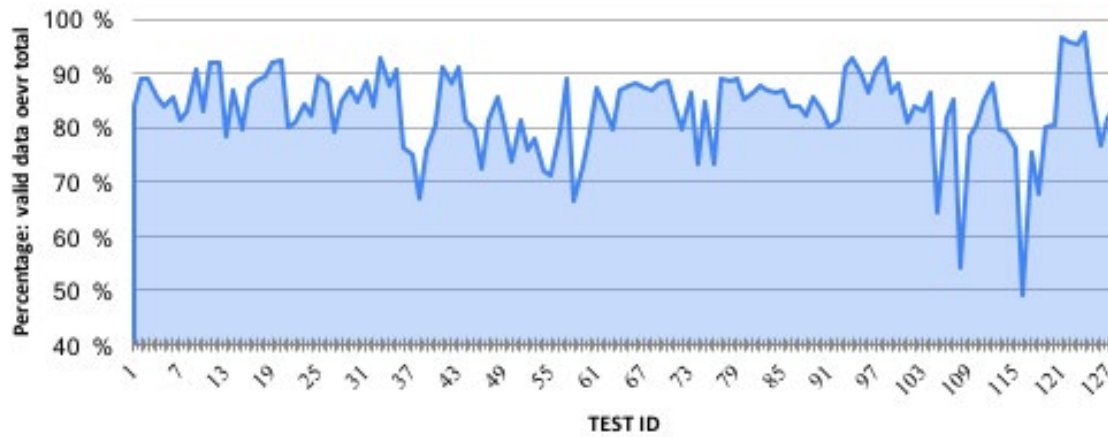
Figure 17. Distribution of Valid Data Over Total Percentage Across the 128 Test Repetitions

IV. RESULTS

The authors divide the discussion of results into three separate sections: (i) the analysis of the suburban scenario; (ii) the analysis of the highway takeover scenario; and (iii) human factors results related to the subjective assessment that each participant provided through the surveys before, during, and after the test.

MANUAL DRIVING AND SUBURBAN SCENARIO

Within the suburban scenario, participants were asked to drive in a residential environment for a target time of two minutes following GPS instructions. After two minutes, a non-driving task related to sending a text message was introduced in conjunction with a traffic light turning red. While replying to the text, the participant needed to stop at the red light behind another vehicle; subsequently, after the light turned green, the participant needed to resume driving while a second text message was received and needed a reply. The first text message received by the drivers asked the drivers to provide their estimated time of arrival. It was then followed by a second message asking them whether they preferred a cheese or pepperoni pizza for dinner. For both messages, participants were instructed before the beginning of the test to reply when comfortable through the specific interface provided (i.e., either manual or vocal). For each test repetition, the participant was told which interface they would be using before the driving began and asked if comfortable with it. Each participant for the 32 completed tests experienced both interfaces and thus repeated the suburban scenario twice. The order of presentation of the interface was randomized to reduce the effect of learning, with a balanced design among the four scenarios tested (manual+suburban; vocal+suburban; manual+highway; vocal+highway).

Figure 18 shows a scheme of the intersection that the drivers encountered and their estimated location when each of the texts were received. The distance from the first text reception point and the traffic light was on average 55 meters, and the distance from the second text reception point and the traffic light was on average 8.3 meters.
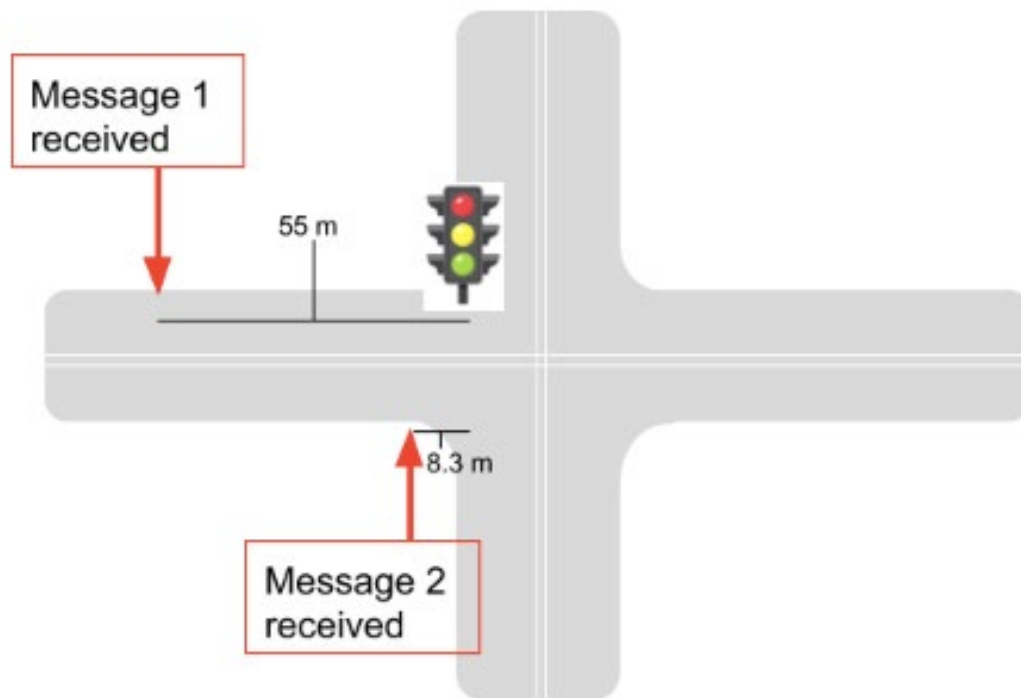
Figure 18. Intersection and the Text Messages Locations: Suburban Scenario

Analysis of Time-to-Throttle

The first dependent variable analyzed for the suburban scenario was the time-to-throttle. This quantity, measured in seconds, provides the duration of time between the timestamp associated to the traffic light color switching from red to green and the timestamp associated to the first consistent throttle input from the driver. A consistent input was one associated with a clear response from the driver in applying pressure to the throttle pedal, and did not account for spurious recorded vibrations in the throttle input: a deeper discussion on this point is provided within the highway scenario analysis.

Figure 19 provides the graph of the probability density function for the 64 measured time-to-throttle (32 participants for two interfaces). As can be gathered by looking at the distributions, similar behavior was observed for both interfaces: a main peak is observed between one and two seconds, with small variation overall.
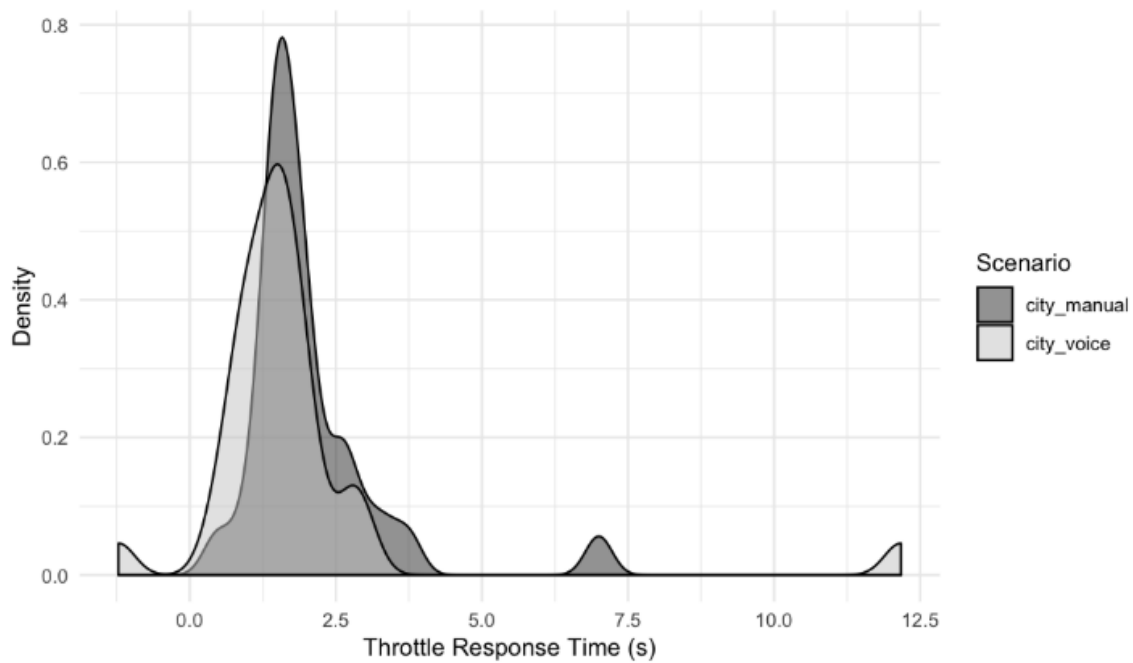
Figure 19. Probability Density Function for Time-to-Throttle: Suburban Scenario

Table 2 provides a summary of the average computed time-to-throttle and standard deviation. A two-tailed t-test was performed, and the difference in values for the two interfaces was found to be not statistically significant.

Table 2.   Summary of Recorded Time-to-Throttle and Pertinent Statistics

Interface	Mean [s]	SD [s]
Manual	2.02	1.15
Vocal	1.77	2.08
Statistics	p = 0.559	NOT SIGNIFICANT

The p-value for the time-to-throttle was 0.559, above 0.05 (the significance level with a confidence interval of 95%). This means that the experiment did not show a statistical difference between the two interfaces with respect to how quickly a person resumed driving from a stopped condition after the external stimulus of a traffic light turning green, when the light switched color at the same time as an incoming text needing reply.

Analysis of Stopping Distance

In order to best understand the drivers' behavior after the first message was received, a plot similar to the one shown in Figure 20 was generated for all the drivers for both manual and vocal interfaces.
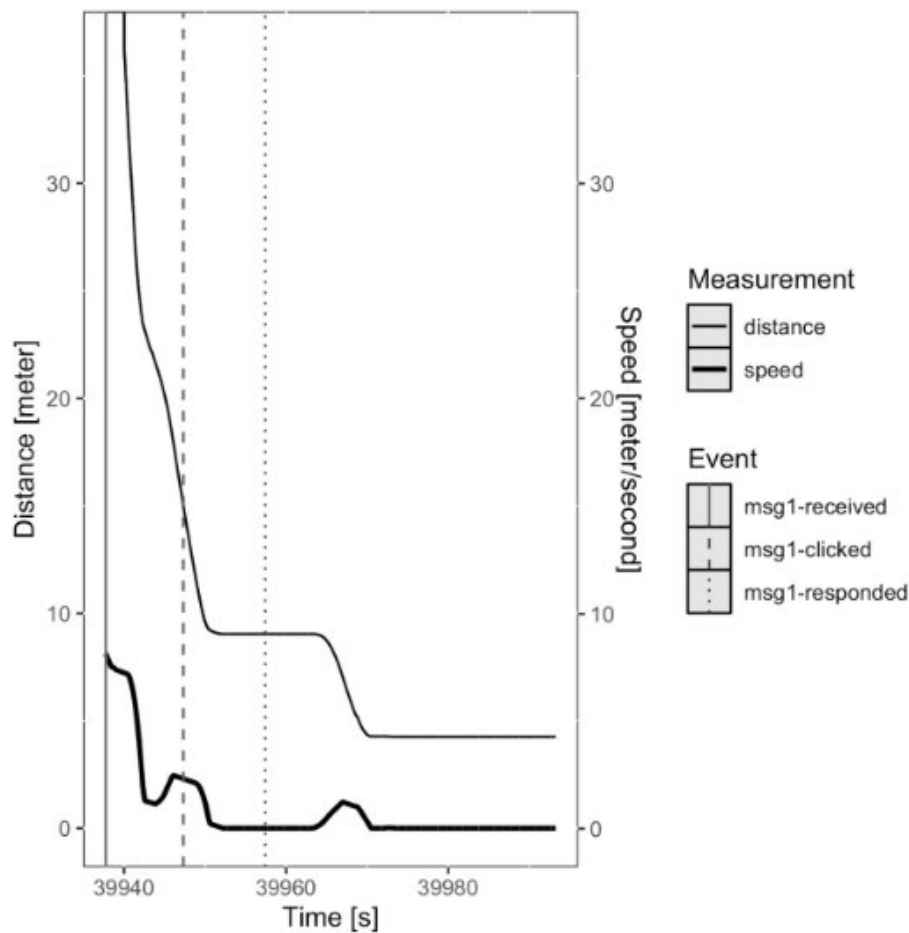
Figure 20. Speed and Distance as a Function of Time for Trajectory Analysis: Suburban Scenario

The plots provided information about the speed of the vehicle and the distance between the ego vehicle (i.e., the study team's vehicle, driven by the participant) and the vehicle ahead, stopped at the red light. The speed of the vehicle is the bolded line on the bottom, and since the driver was reaching the intersection when the traffic light is red, the speed was decreasing among all drivers, eventually going to zero. However, the "bumps" and changes in their speeding behavior up to the final stop are what differentiated the drivers. Some drivers showed a gradual and gentle decrease, while others had many "stop-and-go" transitions, depending on their interactions with the phone and the texting task. Those bumps in speed, and the corresponding flat regions at zero speed, are further correlated to the line on the top, representing the distance between the two vehicles. Such distance remains constant whenever the speed of the ego vehicle is zero. Furthermore, vertical lines in the plot show the times at which the first text message was received (solid line), the driver first clicked on the phone (dashed line), and the reply message was sent (dotted line).

As an example, the driver in Figure 20 slowed down in two steps from 8.1 m/s to 1.1 m/s and then increased to 2.4 m/s before touching the phone to read the text message, came to a stop to send a response at a 9-meter distance from the car ahead, and then sped to 1.2 m/s to finally stop at the traffic light with the two vehicles apart by 4.3 meters. While the

analysis of texting behavior is carried out in the next section, the researchers focus here on the stopping distance to understand whether any difference in the margin left between the two vehicles was observed.

Figure 21 shows the view of the intersection and of the vehicle stopped in front of the ego vehicle from the perspective of the participant.

Figure 21. View from the Participant's Seat of the Vehicle in Front and the Stop at the Traffic Light: Suburban Scenario

Table 3 provides a summary of the statistics for stopping distance. Also, in this case, the difference in measurements between the two interfaces was shown not to be statistically significant. No collision was ever experienced with the vehicle in front, and the minimum recorded distance was 1.4 meters. Considering an average speed of 2 m/s approaching the intersection, the mean distance between the two vehicles left a lapse time of roughly 3.5 seconds (7 m divided by 2 m/s) between the two vehicles. Lapse times ranging from 2 to 5 seconds are commonly employed for adaptive cruise control usage and are considered standard thresholds in highway environments (thus at much higher speeds than what considered here). Based on this consideration, participants in this experiment were considered to be well within appropriate safety thresholds for both interfaces.

Table 3.   Summary of Recorded Stopping Distance and Pertinent Statistics

Interface	Mean [m]	Max/Min [m]
Manual	7.2	11.7/2.3
Vocal	7.0	14.0/1.4
Statistics	p = 0.75	NOT SIGNIFICANT

Analysis of Texting While Driving Behavior

Based on the vertical timestamps indicated in Figure 20, an analysis of texting while driving behavior was carried out. The intention was to generalize behaviors for speed and distance handling during a secondary non-driving task for the two interfaces.

Three categories of behaviors were identified after analyzing all data for both interfaces:

1. **Full stop to respond:** The drivers in this category fully stopped at the traffic light before attempting to respond to the texts received. They remained stopped until the light became green, after which they resumed driving.

2. **Response while changing speed:** The drivers in this category increased or decreased their speed while responding to the texts received before coming to a full stop at the intersection.

3. **No response:** The drivers in this category did not reply to the text. While they were instructed to "reply when comfortable," not all replied to all the texts.

Note that the second text message was received at an intersection right before the light turned green. Therefore, it was of interest to this research undertaking to also study how many drivers actually responded while they were crossing the intersection, which is a fourth added category of interest. Table 4 provides a breakdown of the observed behaviors, where the "response at intersection" accounted for cell phone interactions within 30 meters of the stopped position at the traffic light.

Table 4.   Texting while Driving Behavior: Summary and Breakdown of Observed Behavior

Response Category	Manual, message1	Vocal, message1	Manual, message2	Vocal, message2
Full stop to respond	18 (56%)	22 (69%)	6 (19%)	6 (19%)
Response while changing speed	14 (44%)	9 (28%)	24 (75%)	23 (72%)
No response	0 (0%)	1 (3%)	2 (6%)	3 (9%)
Total number of tests	32	32	32	32
Response at intersection	NA	NA	20 (66.7%)	14 (47.6%)

Various interesting observations could be made based on the data in Table 4. In particular, the analysis of the first text highlighted the following behaviors:

- With respect to the full stop before responding behavior, within the manual interface, it was shown that on average, drivers waited 1.6 seconds after coming to a stop before engaging with the phone. In manual mode, the longest recorded wait time was 10.5 seconds, and the minimum wait time was 0.1 seconds after full stop. For the vocal interface, the average wait time increased to 3.9 seconds, with the longest time being 17.2 seconds, and the lowest 0.05 seconds.

- With respect to recorded responses with a moving vehicle, the manual interface wages an average speed during the first phone interaction of 4.1 m/s (9.2 mph), with the highest recorded speed being 10.5 m/s (23.5 mph) and the lowest 1.8 m/s (4.2 mph). For the vocal interface, the average speed during the first phone interaction was 3.1 m/s (6.9 mph), with the highest recorded speed being 6.8 m/s (15.2 mph) and the lowest 0.4 m/s (0.9 mph).

Overall, participants using the manual interface engaged sooner with the technology, and at higher speeds. Two-tailed t-tests showed marginal significance with p-value at 0.055 for the wait time of engagement (a factor to be analyzed shortly as "time to interact"), and no statistical significance was found for the speed of engagement, with p-value of 0.353

The analysis of the second text also showed similar conclusions, with quantification as follows:

- With respect to the full stop before responding behavior, within the manual interface it was shown that on average, drivers waited 14.6 seconds after coming to a stop before engaging with the phone. In manual mode, the longest recorded wait time was 33.2 seconds, and the minimum wait time was 0.5 seconds after full stop. For the vocal interface, the average wait time increased to 27.5 seconds, with the longest time being 40.4 seconds, and the lowest 4.3 seconds.

- With respect to recorded responses with a moving vehicle, the manual interface yielded an average speed during first phone interaction of 6.0 m/s (13.4 mph), with the highest recorded speed being 12.9 m/s (28.5 mph) and the lowest being 0.1 m/s (0.2 mph). For the vocal interface, the average speed during the first phone interaction was 7.5 m/s (16.8 mph), with the highest recorded speed being 14.1 m/s (31.5 mph) and the lowest close to zero.

The longest wait time for engagement with the second text is explained by reference to the intersections: drivers were less likely to reply to the text while crossing the four-way intersections. Furthermore, note that the trend for speed while engaging with the phone is reversed for the second text, with slightly higher speeds for the vocal interface. Two-tailed t-tests found no statistically significant difference between the two interfaces for wait time or for speed of engagement with cell phone. No other study in the literature was found to analyze the speed at which the drivers engaged with the phone and/or wait times, so a comparison with other data is not possible at this time. The authors recommend that the community look into such quantities which are safety-critical in their aim of ensuring the avoidance of erratic driving behavior when secondary tasks are introduced.

Furthermore, the authors carried out an analysis on texting behavior using metrics such as time to first click, time for composition, and time for overall reply. The following variables were thus defined and checked for statistical significance:

- Time to Interact: time duration from when the text message arrived until the participant made their first interaction with the cell phone;

- Time to Compose: time duration from the first interaction until the participant was finished composing their message;

- Time to Send: time duration from receiving the text message until the reply was sent (the sum of the previous two variables).

Statistical significance was found for both time to interact and time to compose. In particular,

a marginal p-value of 0.052 was recorded for time to interact, with participants prone to click on the phone sooner with the manual interface. Moreover, and intuitively, a p-value less than 0.001 was found for time to compose the message, with participants being faster at composing a vocal message rather than typing (note that the distributions on number of characters contained in the text were checked and were found comparable). Figure 22 shows the boxplots obtained for the two statistically significant variables.
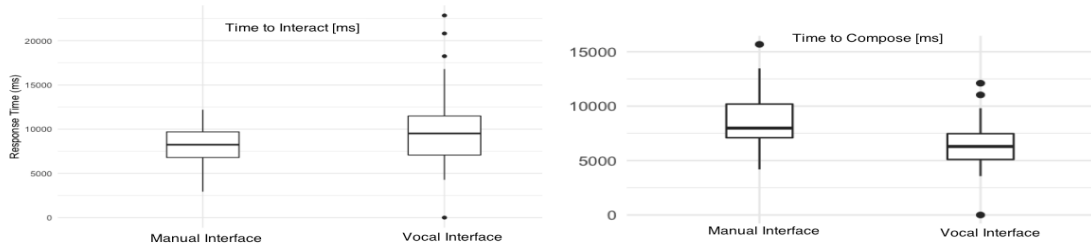
Figure 22. Boxplot for Significant Variable of Time to Interact and Time to Compose: Suburban Scenario

Analysis of Eye Tracking

Two quantities were analyzed with respect to eye tracking:

- The maximum dwelling time: defined as the longest time interval, measured in seconds, that a participant spent continuously looking at one of the six areas of Figure 15 (e.g., windshield, or cell phone).

- The classified time percentage: defined as the percentage of time, from 0% to 100%, spent looking at each of the six zones of Figure 15 over the duration of the test.

Table 5 presents the data for maximum dwelling on cell phone and maximum dwelling on windshield for both the manual and the vocal interface.

Table 5.   Summary of Maximum Dwelling Times: Suburban Scenario

Interface	Class	Mean [s]	SD [s]	Max/Min [s]
Manual	Phone	0.35	0.33	1.32/0.03
	Windshield	4.06	2.19	10.0/0.52
Vocal	Phone	0.27	0.24	0.88/0.02
	Windshield	4.28	2.45	11.08/0.57
Statistics Phone	p = 0.316	NOT SIGNIFICANT		
Statistics Windshield	p = 0.707	NOT SIGNIFICANT		

As gathered from Table 5, dwelling times on the cell phone were found to be lower for the vocal interface, but the change was not enough to show statistical significance. Conversely,

dwelling times on the windshield were longer for the vocal interface, but once again no statistical significance was observed.

Finally, Table 6 presents the data for overall percentage of time on cell phone and windshield. Note that the percentages do not add to 100% simply because only two of the six categories are here analyzed. Also in this case, the difference was found not to be statistically significant.

Table 6.   Summary of Classified Time Percentages: Suburban Scenario

Interface	Class	Mean [%]	SD [%]	Max/Min [%]
Manual	Phone	2	2	5/0
	Windshield	43	23	76/5*
Vocal	Phone	1	1	6/0
	Windshield	45	23	82/5*
Statistics Phone	p = 0.508	NOT SIGNIFICANT		
Statistics Windshield	p = 0.684	NOT SIGNIFICANT		

* Minima obtained in tests with low accuracy of valid data, per Figure 17

HIGHWAY SCENARIO: DISENGAGEMENT TAKEOVER

The analysis of the highway scenario revolves around the control takeover following the disengagement of the autonomous mode. As mentioned in the methodology section, drivers received a 15-second stage alert (or pre-warning) that the disengagement was about to happen. Drivers were also told before the beginning of the test that the vehicle was not capable of maintaining the autonomous mode in a highway exit, and so they would be prompted to regain control of the vehicle if the GPS indicated that an exit was to be taken. During the staged alert phase, the drivers would be engaged by a secondary non-driving task: a text message prompting them to provide their estimated time of arrival.

Analysis of Time-to-Input Variables

As for the suburban environment, the first dependent variables to be analyzed were time-related. While in the suburban scenario the researchers only focused on time-to-throttle (given that the car was starting from a stopped condition on a straight road), the highway analysis accounted for all three possible inputs the driver could provide: steering, throttle, and braking. In the highway scenario it was important to analyze all three variables, given that the car was traveling at a speed of roughly 55 mph, and the driver was required to change lanes and take an upslope exit.

Similar to the problems encountered in a 2018 MTI report by Favarò and colleagues, in order to select the response time (whether in relation to steering, throttle/acceleration, or brake usage), it is important to understand what is considered to be the first consistent response. In other words, it is important to pinpoint the exact time that corresponds to a deliberate action of the driver in order to execute a specific maneuver (i.e., steering or

pushing one of the pedals).

What the authors observed was that a simple threshold "!0" (different from zero) would not work in most cases, as the computer would automatically select unreasonably small reaction times (on the order of 1 ms) due to vibrations in the vehicle, or due to small, non-deliberate movements of drivers who were resting their hands on the wheel for comfort, thus reflecting an involuntary action. Moreover, the automation had a lag time of about 10 ms in shutting off all automated outputs.

In order to more accurately select the correct response time, the researchers proceeded to create a visual method that consisted of plotting the response logged by the computer and finding the foot of the peak of the first consistent action, i.e., an action that was aimed at either: (i) steering the vehicle in the correct direction following the road; (ii) accelerating the vehicle; or (iii) decelerating the vehicle. Visually, this is illustrated in Figure 23.
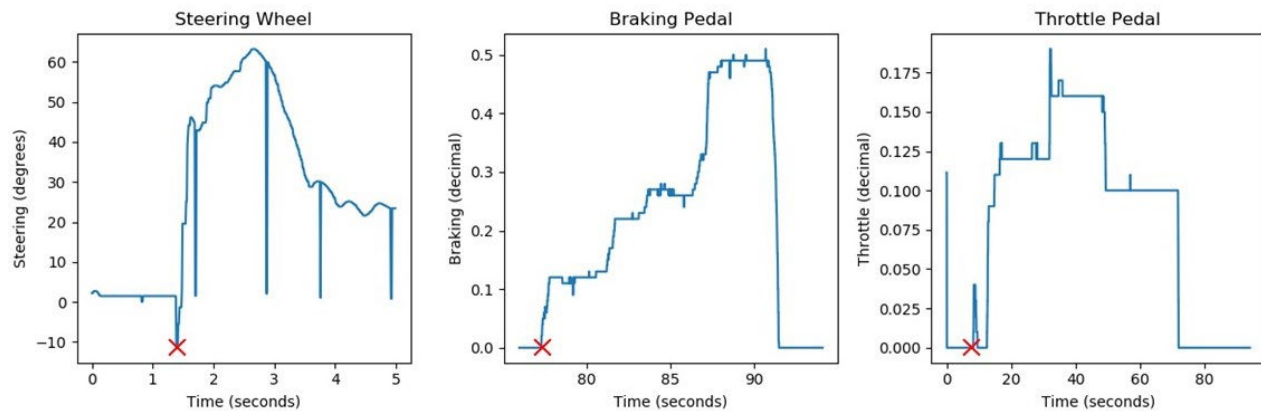
Figure 23.   Representation of the Visual Method for Response Time Selection

The method employed is best captured in the leftmost graph of Figure 23. The plot represents the steering wheel angle captured by the simulator computer. As can be seen from the leftmost picture, the majority of the recorded responses are on the positive side of the abscissa, but two small negative peaks are also recorded before what the researchers call "the consistent response." The actual recorded response time used in this work is indicated by the red mark, placed at the foot of the positive peak. This is the time at which the researchers first recorded a response from the driver to steer the vehicle in the correct direction. Note that in no case did the researchers observe a consistent response in the wrong direction. The researchers carried out the same process for braking and throttle pedal usage. In those cases, the selection was easier, as only positive values can be recorded, and the algorithm developed only had to find reasonable peaks (i.e., those that removed small vibrations that were clearly traceable to computer errors or involuntary actions).

Table 7 provides a summary of the different time-to-input values observed in the study, with means and standard deviations for both interfaces.

Table 7.   Summary of Collected Time-to-Input: Highway Scenario

Input	Manual Mean [s]	Manual SD [s]	Vocal Mean [s]	Vocal SD [s]
Throttle	1.90	1.78	3.80	7.87
Steering	0.77	0.41	0.74	0.33
Braking	18.21	7.75	16.93	6.48
First Response	0.57	0.38	0.55	0.41
Statistics Throttle	p = 0.194	NOT SIGNIFICANT		
Statistics Steering	p = 0.714	NOT SIGNIFICANT		
Statistics Braking	p = 0.480	NOT SIGNIFICANT		
Statistics First Response	p = 0.819	NOT SIGNIFICANT		

Note that in addition to the three inputs, the researchers calculated a fourth variable designated as "first response," which indicated the time-to-first-input used, whichever one it was for each participant. In 62.5% of the cases, the first input was steering, while the remaining 37.5% of cases showed throttle usage as the first input. None of the variables were found to show a statistically significant difference for the two interfaces.

Analysis of Lateral Offset

In order to measure the quality of the control takeover, the authors examined, for each driver, drift from the lane center line, and the authors then compared the level of drift obtained during takeover following system disengagement to the baseline level of drift obtained in the manual driving practice phase.

The analysis focused on two main metrics for the quantification of the takeover performance. The first metric is the lateral offset after the takeover request (TOR) up until the vehicle merged within the exit lane (over a distance of approximately 350 meters). Peak offset, average offset, and standard deviation were analyzed to capture the drift behavior of the vehicle following the disengagement. The second metric, which the authors call integral offset, encompasses the overall behavior within the selected 350 meters (not just the peak) by computing the integral of the car's lateral offset from the lane center line. The integral offset so obtained during the test is then compared to the integral offset obtained during the manual training portion of the test, in which drivers executed an exit-lane merger after a 5-minute manual drive within the simulator. Note that the manual training always occurred prior to the initiation of the automated test. This integral offset ratio allows manual driving performance, for every participant, to act as a performance baseline against which to assess their simulated driving performance.

Lateral Offset

Figures 24 and 25 summarize the trajectories driven by all participants after resuming manual control of the vehicle while being engaged by the secondary task of a text message awaiting reply.
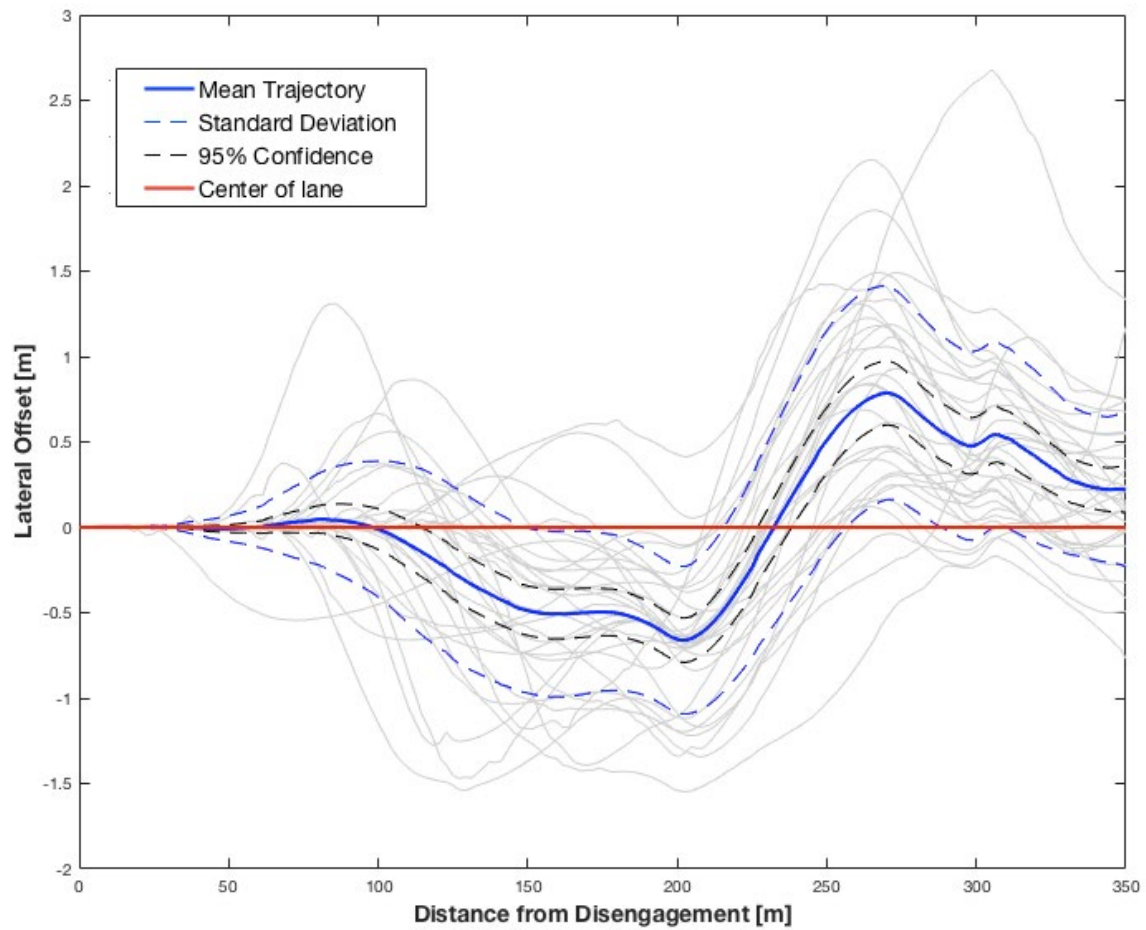
Figure 24. Lateral Offset Summary of Tests for the Manual Interface
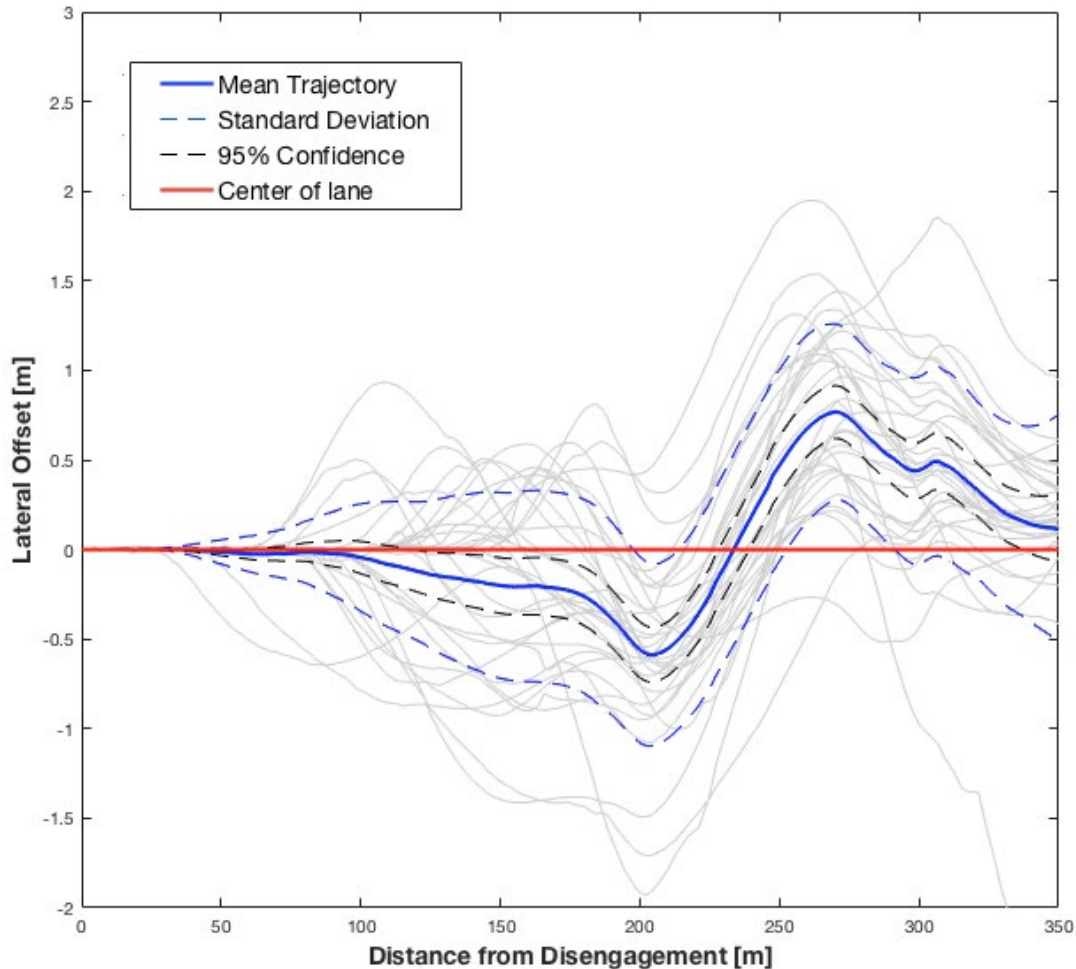
Figure 25. Lateral Offset Summary of Tests for the Vocal Interface

In particular, Figure 24 provides an overview of the trajectories for the tests that employed a manual cell phone interface, while Figure 25 provides an overview of the tests that employed the vocal-reply interface.

The "zero" red line in both figures represents the center of the driving lane, meaning that a perfect trajectory that remains aligned with the lane centerline would appear in both figures as a straight line with a constant zero value. Lane deviations to the right are assigned negative values, while lane deviations to the left are positive. All disengagements happened in the rightmost lane of the three-lane highway environment, with the vehicle subsequently needing to merge into the exit lane on the right. Figures 24 and 25 also depict the respective standard deviations (dashed blue lines) and the 95% confidence interval for the observed trajectories (dashed black lines). Finally, the solid, bold blue trajectory in each plot describes the mean trajectory, while the thin grey lines are the individual trajectories of each participant.

A first important consideration stems from considering unintended and sudden lane departures. The lane width used for this study was 3.6 meters, with a total vehicle width of 1.9 meters. Considering that the center line is located in the middle of the lane, this implies

that a lateral offset greater than 0.85 meters indicates that the vehicle is crossing a lane marking with the outer wheels. The reader can thus imagine two straight lines located at +0.85 and -0.85 for Figures 24 and 25 (not depicted to avoid further clutter); all those trajectories that intersect with either one would thus imply an unintentional lane departure of the ego vehicle. Unintentional lane departures were observed in 71.9% of the cases for the manual interface and in 65.6% of the cases for the vocal interface. It is thus important to determine whether there was a statistically significant difference in the lateral offset behavior across the two interfaces.

Table 8 provides a summary of the data for average lateral offset (across the 350 analyzed meters from the TOR) and maximum lateral offset.

Table 8.   Summary of Data for Lateral Offset

Variable	Manual Mean [m]	Manual SD [m]	Vocal Mean [m]	Vocal SD [m]
Average lateral offset	0.35	0.14	0.31	0.11
Maximum lateral offset	1.22	0.47	1.17	0.59
Statistics average lateral offset	p = 0.219	NOT SIGNIFICANT		
Statistics maximum lateral offset	p = 0.701	NOT SIGNIFICANT		

No statistical significance was found for lateral offset, meaning that the participants performed similarly when replying to a text while taking over control of a vehicle previously driven in automated mode. Moreover, the performance was troublesome, considering the high percentage of unintentional lane departures observed. Visually, the distributions for lateral offset are provided in Figures 26 and 27, where it is possible to intuitively learn of the missed statistical difference indicated by Table 8. In other words, the p-values indicated in Table 8 are indicative of a similarity between the distributions, which is visually easily spotted in Figures 26 and 27.
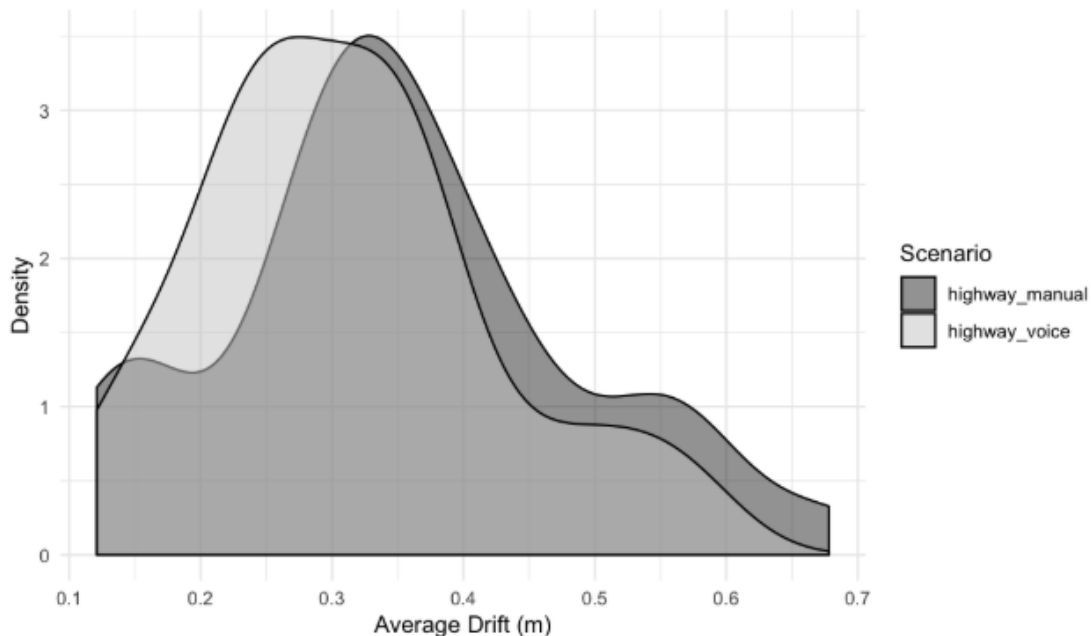
Figure 26. Probability Density Function for Average Lateral Offset
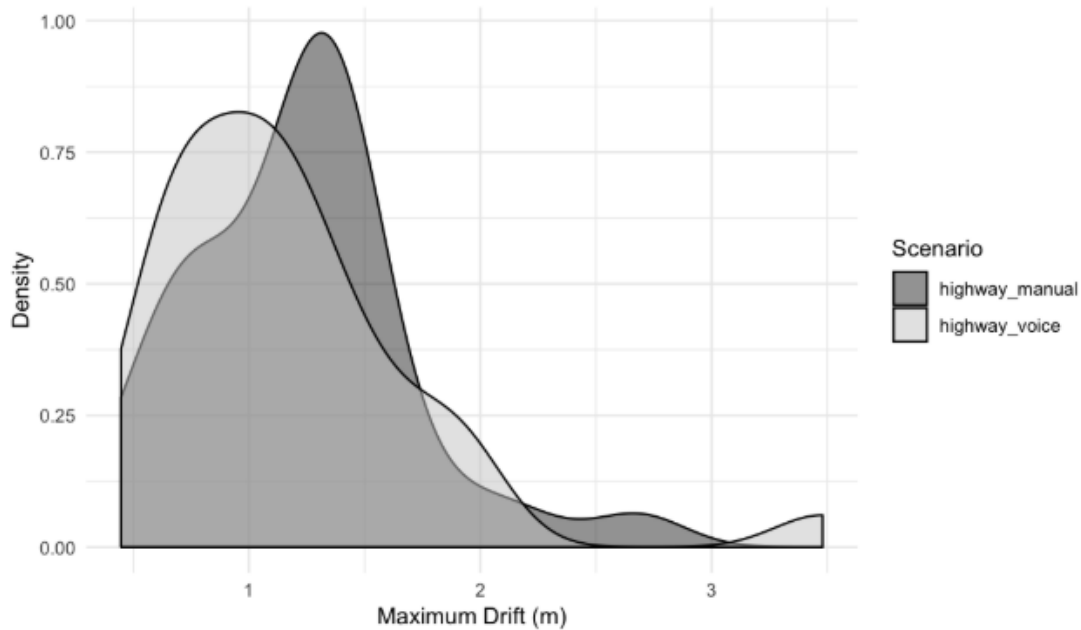
Figure 27. Probability Density Function for Maximum Drift

Integral Offset: Comparison with Conventional Manual Driving

A similar analysis was carried out for the second dependent variable of interest, the integral offset ratio. The goal was to compare the recovery performance of each participant to their own manual baseline. To compute this metric, the researchers first computed integral offsets for both the manual and the automated tests for each driver. The integral offset computes the total area of the curve that each grey trajectory of Figures 24 and 25 forms with respect to the lane center line (i.e., the area between the red line and a thin grey line in Figures 24 and 25). Figure 28 provides an overview of the integral offset distributions for both the manual and automated performance for the 36 tests.

Figure 28. Probability Density Function for the Integral Offset

The two distributions were found to be similar and yet again did not show statistical difference (p-value = 0.218). However, the analysis of the integral offset was important for comparing similarity with the baseline obtained for each driver in conventional manual driving. In other words, each driver was recorded as driving the exact same exit lane merger in conventional manual mode with no distractor. The integral offset ratio was then computed by dividing the integral offset of the automated test (wh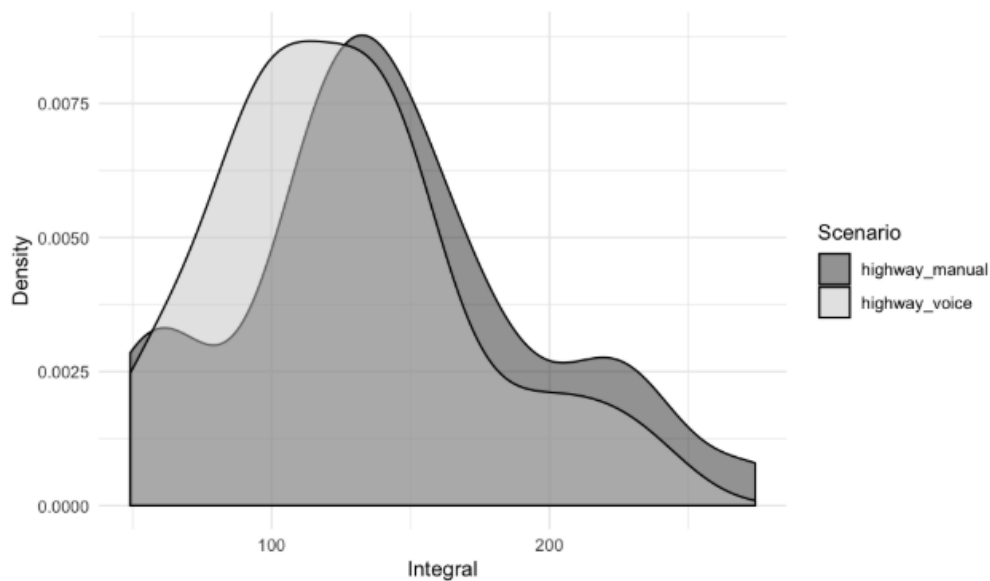ose distribution for the two interfaces is shown in Figure 28) by the integral offset recorded during the baseline manual drive. An integral offset ratio of 1 thus implies an overall similar performance along the first 350 meters of the exit lane merger for drift between conventional driving mode and recovery after disengagement with a secondary task. Values higher than 1 signify that performance is worse during recovery after disengagement compared to during conventional driving; these are the values that the researchers expected to see. Values lower than 1 signify that participants performed better after recovery from a system's failure than during conventional driving modes. The dimensionless ratio is adopted for ease of interpretation instead of a difference, which would have square meters (or square feet) units.

Table 9 provides a summary of the data for the integral ratio. The variable, also found not to be statistically significant, indicates that the interface modes for cell phone interaction investigated in this experiment showed no statistical difference in their comparison between manual driving and control takeover after disengagement in the presence of a secondary task. In other words, participants on average worse did worse compared to manual driving, with no statistically significant improvement in drift or lane keeping when a vocal interface was used over a manual one. Furthermore, note that some participants were not affected by the disengagement and the text reception/reply process to the same extent. In the manual text reply case, 7 out of 32 (21.8%) participants performed better than their corresponding manual baseline drive; in the vocal reply, 11 out of 32 (34.4%) participants showed an integral ratio lower than one.

Table 9.   Summary of Data for Integral Offset Ratio

Interface	Mean	SD	Max / Min	Counter of tests with ratio <1
Manual	1.79	0.92	3.91 / 0.24	7
Vocal	1.57	0.80	3.99 / 0.51	11
Statistics integral offset ratio	p = 0.300	NOT SIGNIFICANT		

Analysis of Texting Behavior

The texting behavior analysis for the highway scenario focused on analyzing the time to interact, time to compose, and time for overall response. In this scenario, it was not possible to categorize and correlate speed and texting behavior, given that the car was not starting from a stopped position and that participants were unlikely to stop to reply to a text in a highway exit environment (which, in fact, none of them did).

Similar to the suburban scenario, statistical significance was shown for the time to compose, with lower composition times obtained in vocal replies. Also, in this case the number of characters used for both the manual and vocal responses were comparable. Figure 29 shows the boxplot for the time-to-compose distribution for the two interfaces (p-value = 0.047).
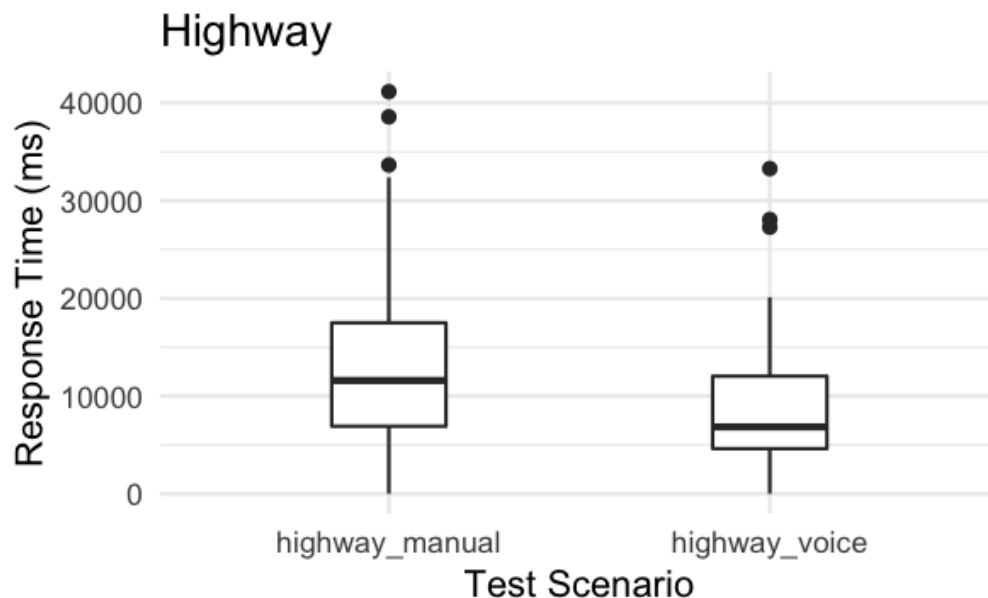
Figure 29. Boxplot for Time to Compose: Highway Scenario

Analysis of Eye Tracking

As in the suburban scenario, maximum dwelling time and breakdown of time percentage spent gazing at various areas of interest were analyzed for the highway scenario. Table 10 presents the data for maximum gaze dwelling on cell phone and maximum dwelling on windshield for both manual and vocal interface.

Table 10. Summary of Maximum Gaze Dwelling Times: Highway Scenario

Interface	Class	Mean [s]	SD [s]	Max/Min [s]
Manual	Phone	0.22	0.21	0.75/0.02
	Windshield	3.86	2.85	11.75/0.25
Vocal	Phone	0.14 0.15		0.53/0.02
	Windshield	4.02	2.35	10.70/0.20
Statistics Phone	p = 0.172	NOT SIGNIFICANT		
Statistics Windshield	p = 0.820	NOT SIGNIFICANT		

As gathered from Table 10, gaze dwelling times on the cell phone were found to be lower for the vocal interface, but the change was not enough to show statistical significance. Conversely, dwelling times on the windshield were longer for the vocal interface, but once again no statistical significance was observed. These results echo those presented in the suburban scenario. Furthermore, note that the analysis of gaze for the highway scenario was focused only on the core portion of the test from the time the driver received the pre-warning of imminent disengagement until the end of the test. The data of Table 10 thus capture a portion of automated driving when the participant was "primed" by the staged alert, then received the text, and subsequently had to take over control during autonomy disengagement while merging in the exit lane. The data prior to that (i.e., from the beginning of the test, when the vehicle was driving fully autonomously and the participant was not engaged with any secondary task) are not included, as that portion of the test was only necessary to get to the disengagement, and it did not constitute an important piece of the analysis.

Finally, Table 11 presents the data for overall percentage time gazing on cell phone and windshield. Note that the percentages do not add to 100% simply because only two of the six categories are here analyzed. Also in this case, the difference was found to be not statistically significant.

Table 11. Summary of Classified Time Percentages: Highway Scenario

Interface	Class	Mean [%]	SD [%]	Max/Min [%]
Manual	Phone	3	3	11/0
	Windshield	52	28	86/4*
Vocal	Phone	1	2	6/0
	Windshield	49	26	90/1*
Statistics Phone	p = 0.138	NOT SIGNIFICANT		
Statistics Windshield	p = 0.752	NOT SIGNIFICANT		

* Minima obtained in tests with low accuracy of valid data, per Figure 17

SURVEY ANALYSIS: HUMAN FACTORS

In addition to examining the measured engineering quantities, the researchers also

analyzed the subjective feedback from each participant. The researchers asked the participants to complete pre-test, post-test, and in-between-test surveys. This section presents the investigated human factor facets, including recollection of events, workload perceived, estimation of success in various maneuvers, and general human-machine interface (HMI) preferences.

Situational Awareness

Situational awareness was estimated in this study through a number of questions asked in between the scenarios tested as well as after the completion of the overall experiment. The specific areas examined were:

- Recollection of ETA (Estimated Time of Arrival) reported from the GPS;

- Recollection of the speed of the vehicle during the test reported by the digital speedometer;

- Estimation of the time in simulation;

- Recollection of the first input provided to the vehicle after the disengagement;

- Estimation of gaze fixation and dwelling areas prioritized;

- Manipulation check for situational awareness, which checked participants' attention to specific objects placed within the external environment.

In order to assess and interpret situational awareness results, accuracy in the answers was measured. The researchers defined "Accuracy," a performance measure commonly used in Machine Learning for results with binary outcomes.[30] Essentially, accuracy is defined as the ratio of correct answers versus the totality of answers. For each question, categories of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were assessed in order to compute accuracy as

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN). \qquad (1)$$

An answer was categorized as true or false depending on whether the participant's answer matched the actual measurement from the simulation. Thus, a True Positive is defined as when the positive answer from the participant matches a positive measured answer; a False Positive would be when a positive answer from the participant actually corresponds to a negative measurement. The accuracy indicator was used to quantify the quality of participants' recollections of the ETA, the speed of the vehicle in the simulation, and their perception of the success of their control takeover after the disengagement.

ETA Reporting

In the texts received during the simulation, participants were asked to reply to the text by providing their ETA. The ETA of the route was displayed at all times in the GPS interface (as seen in Figure 10) on the smart phone. The paper-and-pen surveys after

each scenario also asked participants to report the ETA (see the full survey in Appendix A). The researchers calculated accuracy for two responses: the response they provided as the reply to the text and the response they provided on the in-between survey after the scenario ended. Both the answers were compared to the actual ETA displayed on the GPS and the overall accuracy was computed, as well as the accuracy for different modes, as seen in Figure 30. Overall ETA estimation accuracy for the test was 31.25%. Higher accuracy in the estimation was obtained through the vocal interface and in city scenarios compared to highways. This finding may be due to the higher physical demand required of the manual reply, as well as the higher concentration needed in the highway scenario to regain control of the vehicle. Marginal statistical significance for two-tailed t-test among the vocal and manual interface was shown (p-value = 0.07) for the survey estimations, but not for the data obtained from texts (p-value = 0.91). This means that there was a difference between what participants replied in the text and what they wrote down after the test. This could have happened for a number of reasons, such as the participants not engaging fully with the simulation and replying a made-up number, while then forcing themselves to recollect the ETA upon formal asking during the in-between survey. In the answers provided after the text completion, the authors observed a marginally significant difference, with better ETA recollection when vocal entries were used.
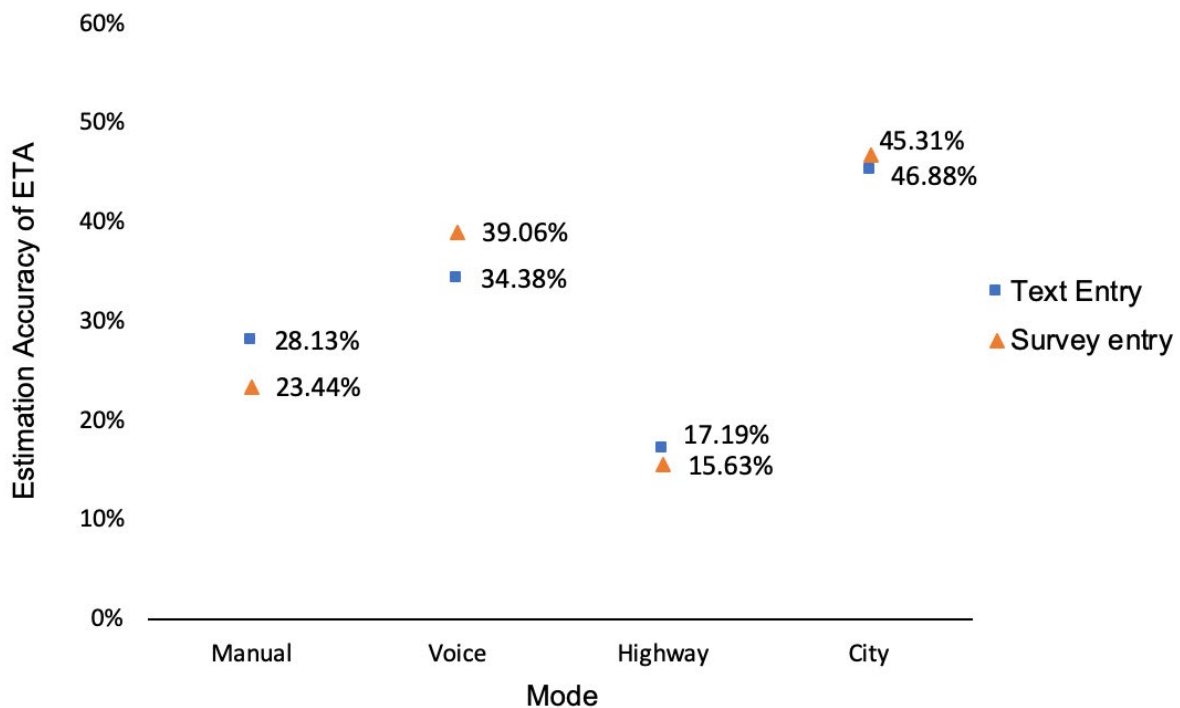
Figure 30. Computed Accuracy for ETA Recollection

Speed Recollection

After each scenario, participants were asked to report the speed at which they were traveling and whether their answer was based on the actual reading of the speedometer. Their responses were evaluated to be correct based on a threshold of ±2 mph with accuracy computed per Equation (1). Overall accuracy for speed recollection was 50%, with individual values for each mode reported in Figure 31. Once more, accuracy for the

vocal interface is higher, and the highway scenario led to a higher accuracy than the city scenario. However, two-tailed t-tests did not show statistical significance for the difference between the two interfaces. The change between highway and city (which was statistically significant, but not among the investigated variables for the study) implies that participants had a better estimate for the scenario of partial automation. Note that the duration of the two scenarios (suburban and highway) was similar.
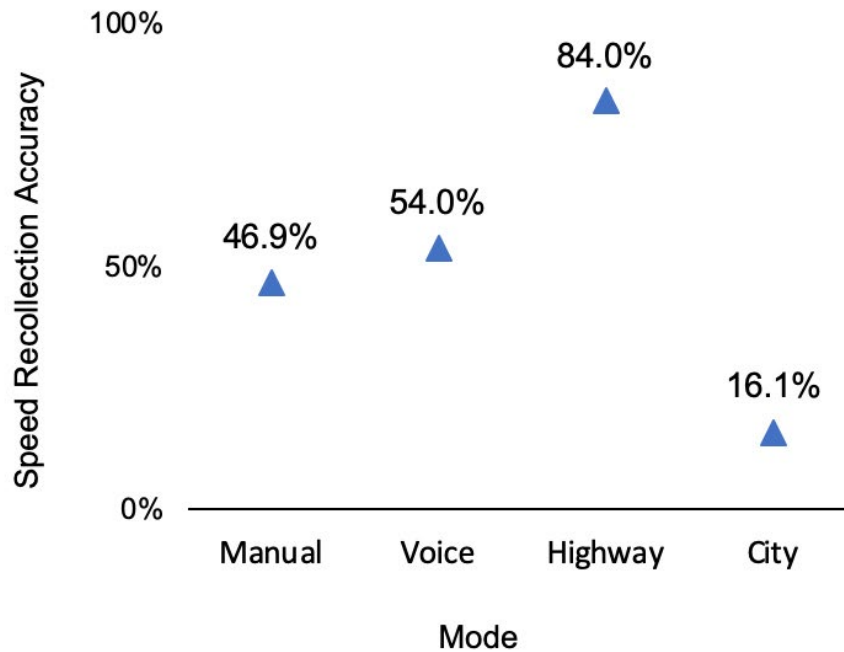

Figure 31. Accuracy of Speed Recollection for Different Scenarios

Estimation of Time in Simulation

Participants were also asked to estimate the amount of time they spent in the simulation. Their reported times were compared to the actual simulation times, which started at the beginning of the practice drive and ended when the last scenario drive was completed. A threshold of ±5 minutes was set to qualify their responses as accurate. Figure 32 shows participants' estimated time versus the actual time spent inside the simulator. Note that only one participant actually accurately estimated the time inside the simulation environment, with the entire rest of the population underestimating the duration. Accuracy was low, with the highest deviation from the correct time being 130 minutes from the actual duration of the test.
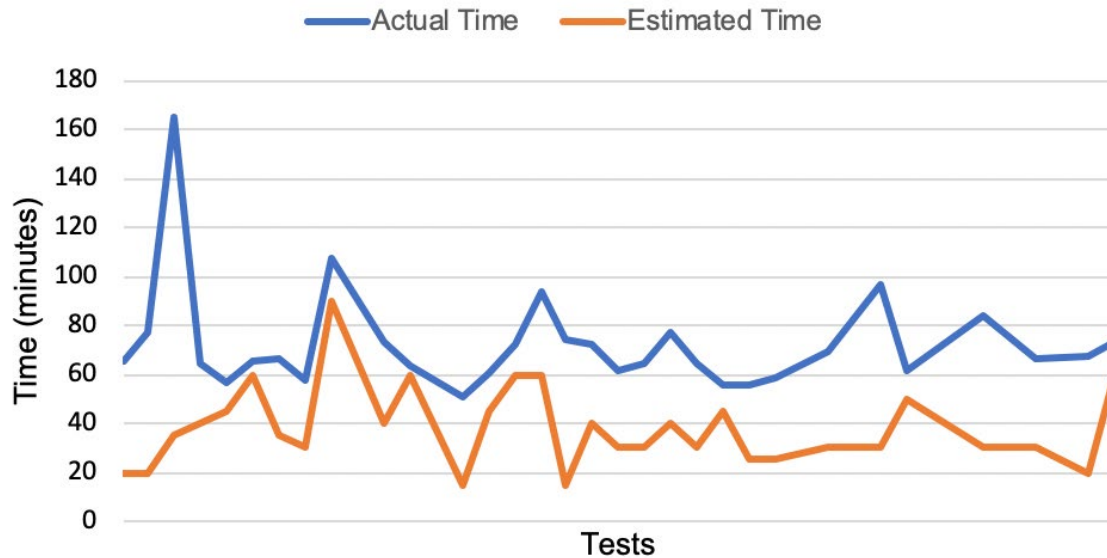
Figure 32. **Comparison of Participants' Estimation of the Time Spent in the Simulator vs. Actual Time in the Simulator**

Recollection of First Input After Disengagement

Participants were asked what their first response was when the disengagement happened in the highway scenarios, in terms of the input used to first control the vehicle motion. Their answers were then compared to their actual first input after the disengagement recorded by the simulator (steering, throttle, and braking). Overall, the responses for the entire population were 63.5% accurate. Seventeen percent of the responses incorrectly stated braking as the first input, whereas braking was not recorded as the first input in any of the cases. Accuracy in responses was a little higher in the vocal mode (67.7%) than the manual mode (59.4%), which could be explained by the manual effort required in responding to the text. Note the similarity with the authors' previous study on disengagement takeover,[31] where participants had also incorrectly recalled braking as the first response, whilst they had instead accelerated.

Estimation of Gaze Focus Areas

Participants were also queried to estimate gaze focus areas for each test repetition in order to compare their responses to the camera data as another situational awareness check, and to check the drivers' perception of fixation on the cell phone. In particular, participants were asked to rank different focus areas when they received the text in both scenarios, and when the takeover request occurred during the highway scenarios. Figure 33 shows the reported focus areas at the time of text reception and at the time of disengagement.
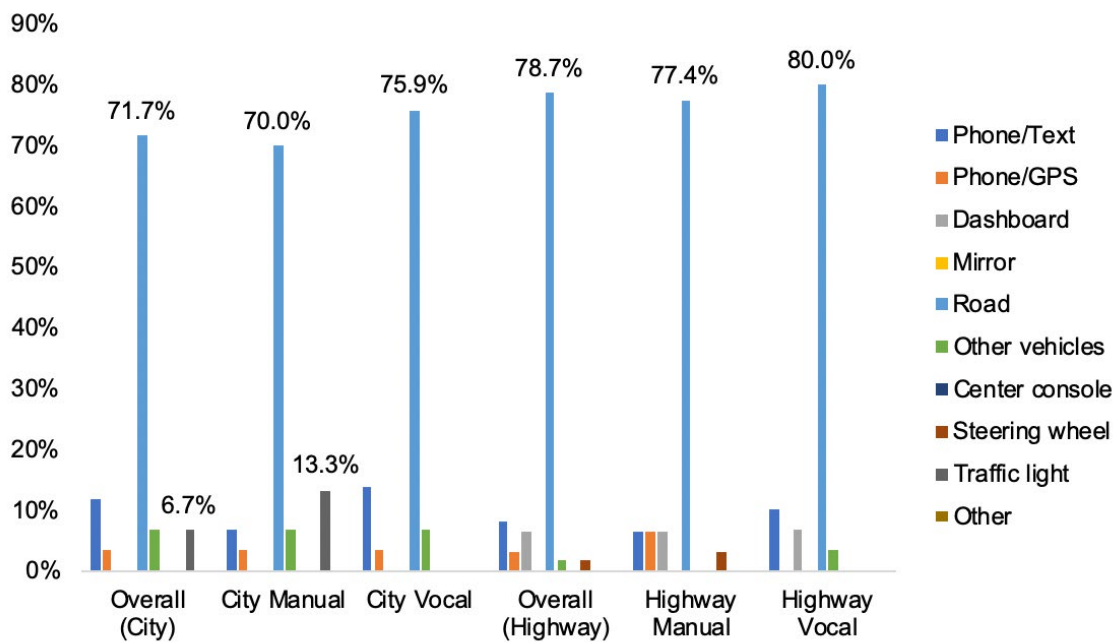
Figure 33. Participant Self-Reported Gaze Levels After Receiving Text Message

The highest-ranked location is the road in all cases, with percentages in the highway scenarios slightly higher than in the suburban (labeled "city" in the figures) scenario. This trend matched what seen from the actual measurements (data for gaze on windshield in highway—Table 11—higher than that in the suburban environment—Table 6). The evaluation is off, though, with the actual data discussed in Tables 6 and 11 only showing percentages between 40 and 50% for gaze on the road. This means that participants overestimated the time actually spent looking at the road.

Figure 34 provides the self-reported data for gaze focus after the takeover request. In this case the attention on the road lowers in favor of an increase for the dashboard, which might be explained by the fact that the visual warning was shown, which required them to move their eyes from the road to the warning.
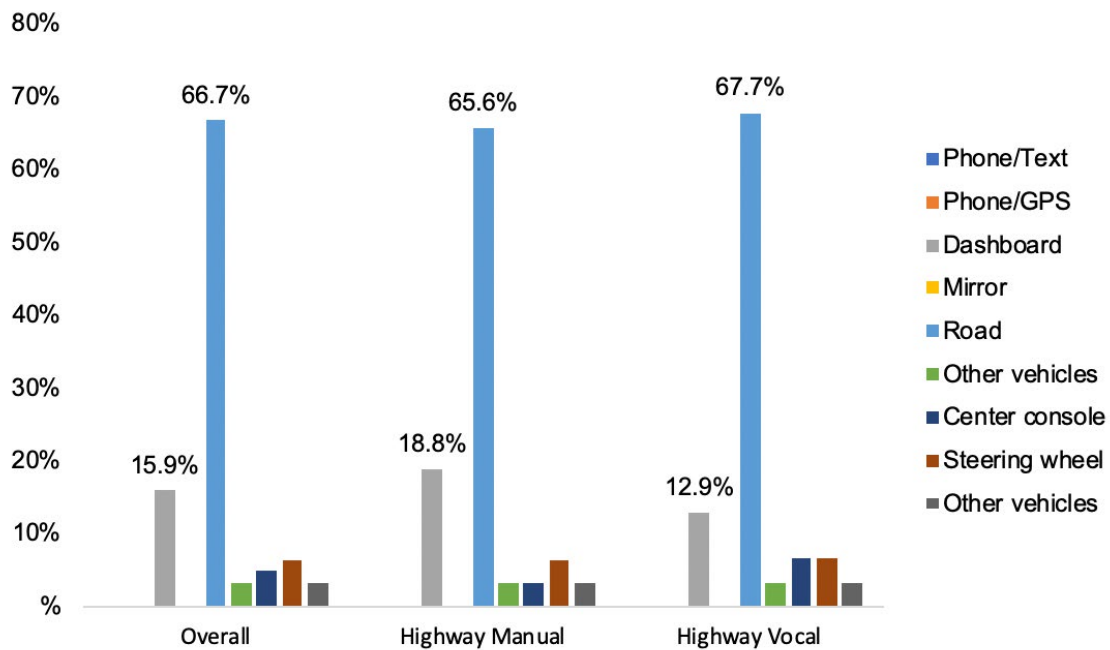
Figure 34. Participant Self-Reported Gaze Levels After Takeover Request

The survey also investigated the strategy employed to reply to the text while remaining mindful of the outside environment. Figure 35 shows that different strategies were adopted for the suburban and the highway scenarios: for the suburban case, over 40% of the participants utilized short alternating gazes between the phone and outside traffic; in the highway environment, the most common option was instead focusing on the outside environment first, and then the phone. This approach was probably due to the higher speed of the vehicle and the denser traffic for the highway environment.
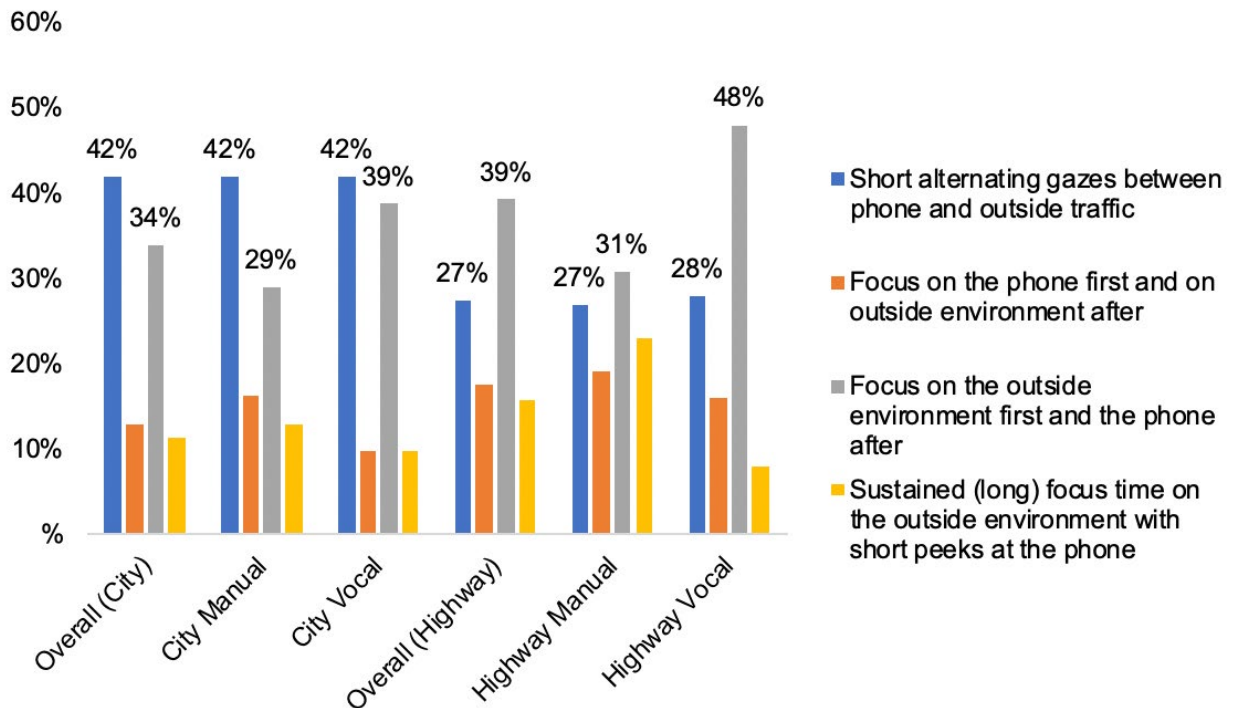
Figure 35. Participant Responses When Asked about the Strategy Employed to Reply to the Texts

Finally, participants rated the gaze fixation ratio on the phone compared to the outside environment. The responses are summarized in Figure 36, showing that the majority of participants indicated that they focused 75% of the time on the environment compared to an estimated 25% on the phone.

Figure 36. Participants' Self-Reported Gaze Fixation Ratio on the Phone Compared to the Environment

Based on the data of Table 6 and 11, participants also overestimated how often they engaged their gaze with the cell phone (overall time spent on the cell phone was rarely over

3–4% of the entire simulation time). In Figure 36, a reported 3% of participants estimated 100% time spent looking at the cell phone for both suburban and highway environments.

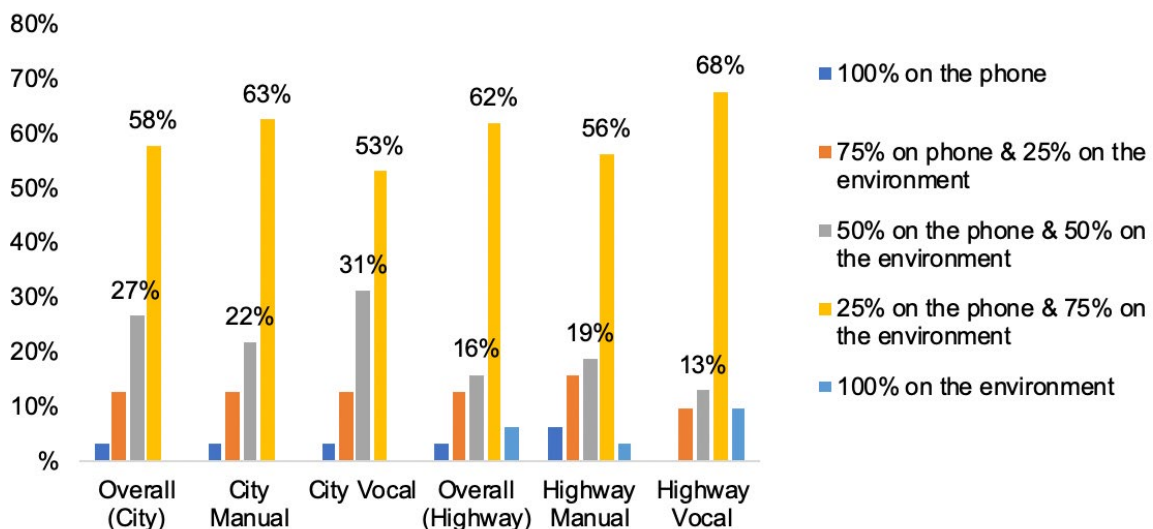The data gathered from the self-assessments was then compared to the data recorded by the SmartEye cameras. Table 12 summarizes the results, which were obtained by parsing the percentages of time spent on each area in chunks of 25% (as in Figure 36) for the competing attention areas of environment and cell phone.

Table 12. Summary of Accuracy for Gaze Estimation

Mode	Matching	Overestimated	Underestimated
Overall	20.51%	56.41%	23.08%
Highway	25.45%	50.91%	23.64%
City	16.13%	61.29%	22.58%
Manual	23.73%	54.24%	22.03%
Vocal	17.24%	58.62%	24.14%

Overall, 20.51% of participants accurately estimated their gaze allocation during the test, with the highest accuracy being in the highway scenarios (25.45%). Note that a slightly higher accuracy was observed for manual modes compared to vocal. It is clear that most people overestimated their gaze on their environment, with more than 50% of participants overestimating the time spent on the environment in all modes and scenarios, the highest being the suburban vocal mode.

Manipulation Check for Situational Awareness

In order to further examine participants' situational awareness, the researchers added a manipulation check similar to the one carried out in the research of Mok and colleagues[32] to determine whether the participants were aware of the external environment. The researchers placed objects along the side of the road that would otherwise be unlikely to be present in the simulation and asked participants if they noticed them. Figures 37 and 38 provide a screenshot of a statue depicting a family of four horses on the shoulders of the highway environment, and a billboard with a giraffe in the front yard of a house for the suburban environment.

Figure 37. Horses Placed for Manipulation Check in Highway Scenario

Figure 38. Giraffe Placed for Manipulation Check in City Scenario

Table 13 summarizes the percentage of participants who were able to spot the objects. The researchers can see a higher accuracy in the highway scenario for the manual interface. Higher percentages for highway scenarios were expected, given the size of the horses and the fact that the vehicle was in automated mode, with participants told to actively monitor the outside environment.

Table 13. Breakdown of Participants Receptive to the Manipulation Check

Mode	Percentage
City Manual	6.25%
City Vocal	12.5%
Highway Manual	59.38%
Highway Vocal	43.75%

Perception of Success of Control Takeover

In highway scenarios, unintentional lane departures were considered as "unsuccessful" control takeovers. The researchers analyzed the drift of the vehicle from the point of disengagement to before the exit (a distance of roughly 350 meters), which was labeled as the takeover period. Success of the maneuver was defined by a lateral drift contained within the bounds of the exit lane (0.85 m). Unintentional lane departure was observed in 68.75% of cases, with the percentage being slightly higher in the manual mode (71.9%) compared to the vocal mode (65.6%). Participants were asked to assess whether their takeover maneuver was successful. Their answers were deemed accurate per Equation 1. Table 14 summarizes the accuracy within their answers. Overall, their answers were 40.6% accurate, with a slightly higher accuracy for the manual mode than the vocal mode.

Table 14. Accuracy of Perception of Success of Control Takeover

| Mode	Manual	Vocal	Overall
TP	9	7	16
TN	3	7	10
FP	18	18	36
FN	2	0	2
Accuracy	43.8%	37.5%	40.6%

Perceived Workload: NASA-TLX

NASA-TLX workload scores from each of the four test conditions were analyzed and compared using effect-size for a repeated measure. Participants were not asked to rate each sub-scale for perceived importance; hence, the researchers use raw TLX scores from the in-between surveys of Appendix A for analysis. The 20-point scale was converted to a measure of percentage, with 20/20 corresponding to 100%. Cohen's effect size is useful for estimating statistical power and sample size, especially if there is low variance. NASA-TLX composite workload scores were produced by combining the six sub-scales using equal weighting.

Drivers' mean subjective ratings for workload are displayed in Table 15 and Figures 39 and 40. The large standard deviations suggest a wide variability in reported individual scores.

Table 15. Descriptive Statistics Summary of TLX Sub-Scores

Mode		Mental	Physical	Temporal	Performance	Effort	Frustration	Total
Cm	M	44.69	42.19	38.44	28.91	49.06	38.91	242.19
	SD	21.32	22.43	21.83	19.54	21.31	23.68	90.29
Cv	M	40.16	36.56	36.56	27.19	40.47	38.13	219.06
	SD	25.19	20.30	24.84	19.05	23.15	23.75	99.85
Hm	M	39.69	26.88	43.44	42.50	41.56	38.91	232.97
	SD	23.21	18.83	27.55	25.08	23.40	25.36	108.90
Hv	M	34.06	25.31	38.44	35.63	39.84	38.13	211.41
	SD	22.30	16.41	24.67	22.10	22.70	25.77	101.14

Notes: city manual = Cm, city voice = Cv, highway manual = Hm, and highway voice = Hv
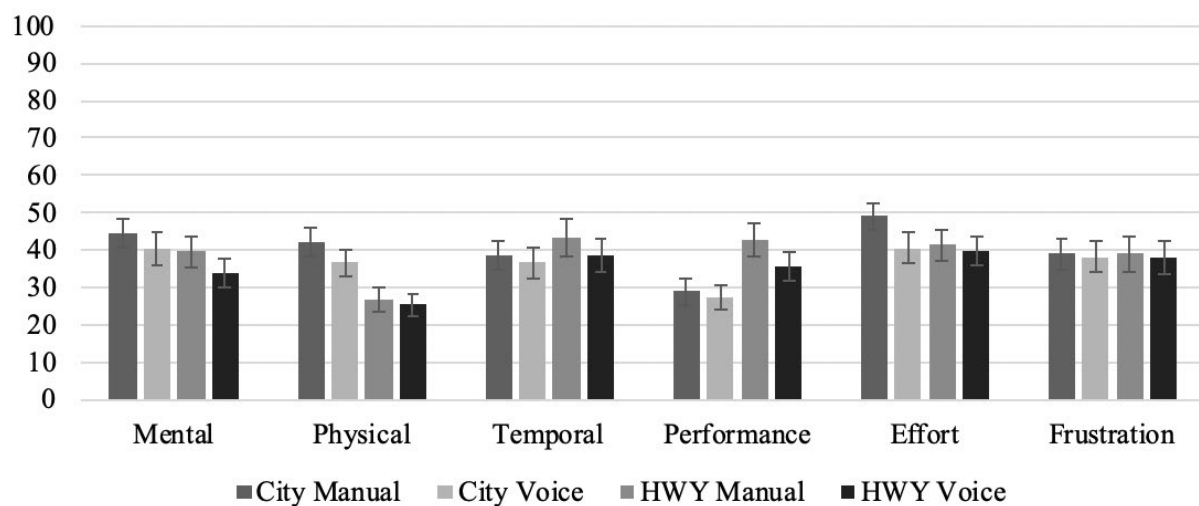
Figure 39. Mean Values for the Sub-Scores of NASA TLX

Looking at the individual sub-scale mean scores across all conditions and participants, workload dimensions with the least differences between all four driving mode conditions were frustration and effort. For the frustration sub-scale, participants' average score in the city manual and highway manual driving modes were identical (M = 38.91), as were the average scores for the city voice and highway voice scenarios (M = 38.13). The dimensions with the greatest differences between driving mode conditions were the physical and performance sub-scales. Mental and temporal dimensions show some variability, as seen in the TLX graph.
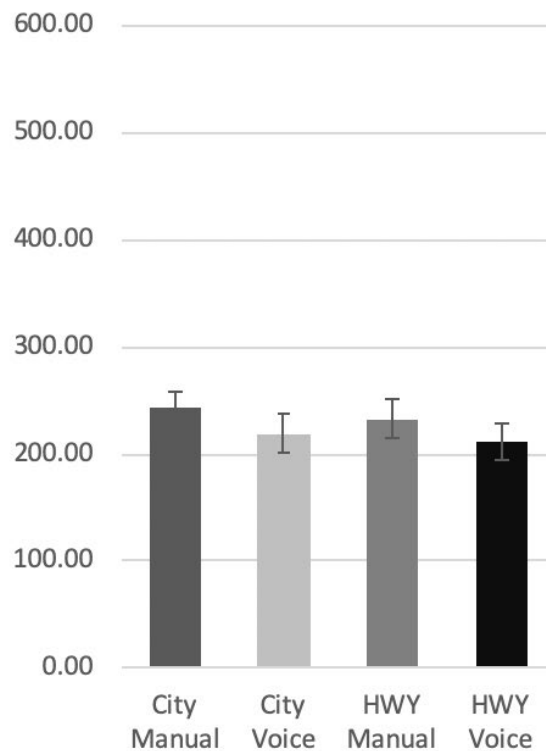
Figure 40. Mean Values for the Global NASA TLX Workload Score Between the Four Driving Mode Conditions

Paired samples within-group comparison t-test analysis revealed that there was a marginal significant difference in participants' perceived mental workload between manual texting (M = 242.19, SD = 90.29) and voice-to-text input (M = 219.06, SD = 99.85) in the city driving mode scenario (p-value = 0.07). No statistically significant difference was found for the highway driving mode scenario during the manual texting condition (M = 232.97, SD = 108.90) or the voice-to-text condition (M = 211.41, SD = 101.14) (p-value = 0.25). When looking at individual sub-scales, users reported higher scores on the effort dimension during the manual text input condition (M = 49.06, SD = 21.31) than during the voice-to-text condition (M = 40.47, SD = 23.15) for the city driving mode scenario, and that difference was statistically significant (p-value = 0.020). No other significant differences were found amongst individual sub-scores in the city and highway mode driving scenarios.

The data suggest that the overall general trend is that voice-to-text technology reduces the cognitive workload of the recruited driver participants in the study regardless of the city or highway driving mode. This is true for all sub-scales and total global scores except for the frustration dimensions. It is possible that perceived feelings of frustration or stress were not a salient aspect of this particular experimental task. The salience of individual scale dimensions for NASA TLX can vary between different tasks, and this claim is consistent with the literature. Although the mental workload was consistently higher when manually interfacing with a cell phone during the secondary texting task, a statistically significant difference was only found on the effort dimension in the city scenario. This sub-scale asks participants to rate how hard they had to work to accomplish the goal, and manual texting added more effort than using voice technology while navigating a city environment

in a simulated manual car. This finding is consistent with the theory of limited attentional resources. Manual texting requires more visual processing, which competes with the primary task of navigating the streets and creates a bottleneck in information processing.

Human Machine Interfaces

Interface Comparison

Participants were asked to specify their preference between the voice and touch interface in terms of personal preference, safety preference, difficulty in use, and estimated promptness of response: whether they thought their reaction to the disengagement in the highway scenario and to the red and green lights in the city scenario was more prompt with one interface or the other. Figure 41 summarizes the responses.
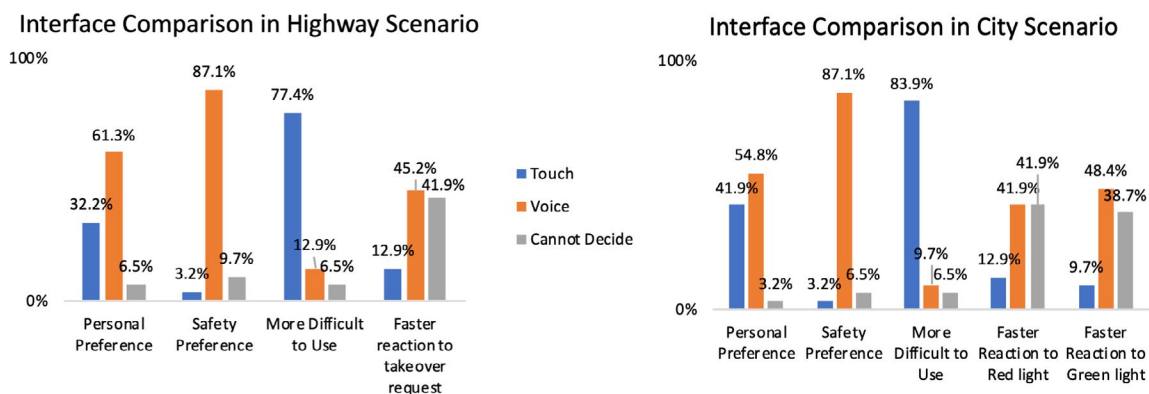

Figure 41. Text Interface Preferences Summary for Smartphone Interface

The majority (77.5%) of participants indicated that they were comfortable with the cell phone interface prior to the test; 6.45% indicated that they were not comfortable (paper instructions as well as video tutorials were provided to the participants ahead of time); the rest of the population was neutral. For both city and highway scenarios, participants showed a slight personal preference for the vocal interface and a significantly higher safety preference for the vocal interface; however, prior to the test, only 12.5% indicated a strong preference for vocal interfaces.

Highway Aural and Visual Disengagement Warnings

The majority of the participants (93%) confirmed they heard the aural warning for the takeover request (TOR). This figure represents a decrease from the author's previous similar study,[33] where all participants heard the aural warning. Two warnings were provided in this test, which prompted the participant to "prepare to take over" and to "take over control" at the beginning and at the end of the "preparation mode" displayed to the right of Figure 5. The majority of participants (80%) also reported the aural warning was helpful and were satisfied by it, but 12.9% found it distracting and one participant said it hindered their ability to take control. With respect to the TOR visual warning and alternation of the three icons of Figure 5, 23% of the participants reported not seeing it within the central

console (10.2-inch display). This is an improved percentage from the authors' previous study (where 50% of people did not see the warning within the same simulator), which could be explained by the fact that the cell phone was placed near the central display in this setup, as seen in Figure 16. Out of the participants who saw it, two reported that it was distracting, but that it did not hindered their ability to regain control, while another participant indicated that it hindered the ability to take over control. A majority of participants (54%) indicated that they were satisfied with the warnings provided, while 6% disagreed.

The researchers also asked participants about their preferences with respect to the warnings. Note that those preferences were not tested: they were simple options participants were asked about during the post-test feedback. The majority (77%) preferred text compared to a light for the visual warning, with more preferring flashing text (51.6%) compared to solid text (25.8%). Red (35.4%) and yellow (32.3%) were the preferred colors for the warning. For the location, the dashboard was the most preferred by a majority (45.1%) of the participants, followed by the HUD/Windshield. This is a slight shift from the authors' previous study, in which the researchers observed that the preference for the HUD/windshield was slightly higher than for the dashboard. Note that 46.8% of the participants mentioned that their personal cell phone is located on the dashboard/vent mount when they drive in real life, which could explain why they would want the visual warning to be placed there, since it might be an easy location for them to glance at. Out of all the text options the researchers provided to express a preference, most participants preferred "Take Control" (48.3%) as the warning message, followed closely by "Disengagement" (38.7%).

The researchers also asked participants about alternative preferences for warnings, and among all options, steering wheel (38.7%) and seat vibration (35.4%) were the most preferred. Among other answers, some participants mentioned that lowering music would be helpful to alert the disengagement in addition to the vibration.

Safety Awareness

Perceived Safety Compromise

Participants were asked to rank on a scale of 1 (being least likely) to 5 (being most likely) whether responding to the texts compromised safety. The results are summarized in Figure 42. The highway scenario showed a higher level of perceived compromised safety, likely due to the increased speed, the increased complexity of the control takeover maneuver, and the surrounding traffic. For both highway and suburban scenarios, manual texting had a higher level of perceived safety compromise. This echoes the results found in the TLX analysis with respect to effort, with participants indicating a higher level of effort required for manual texting and thus a heightened perception of compromised safety.
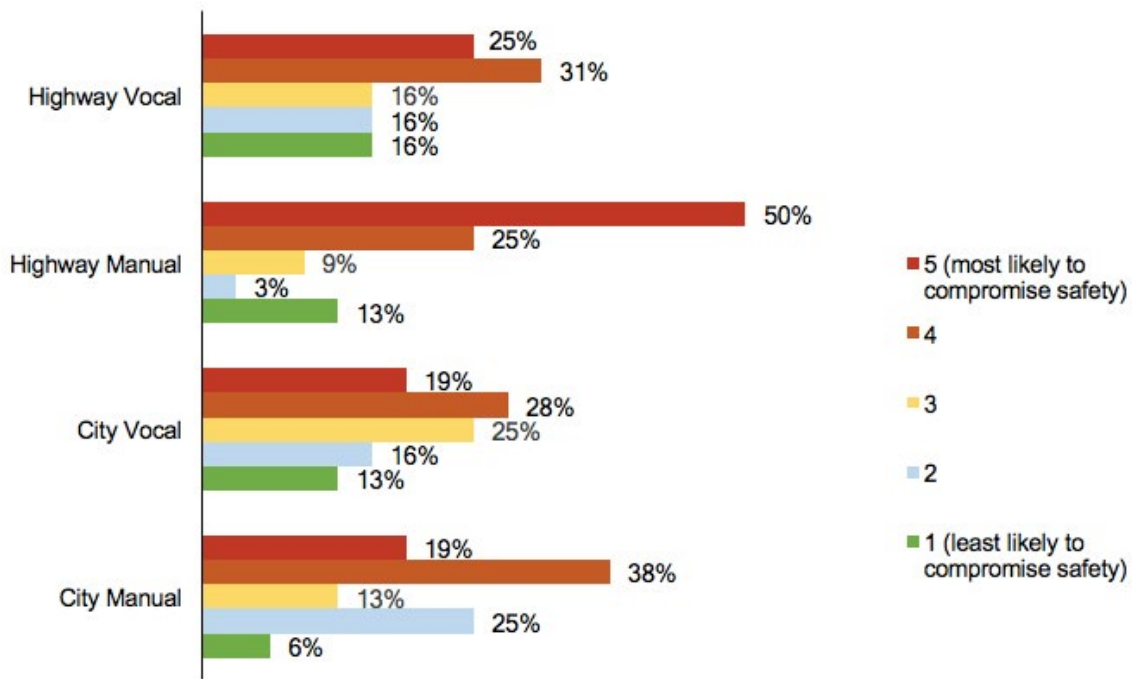
Figure 42. Perceived Safety Compromise as a Function of the Investigated Variables

Subjective Measures

Participants were asked to rank the extent to which they experienced the following emotions before and after the test: tiredness, concern for physical and mental wellbeing, anticipation, anxiety, and fear. Figures 43 and 44 (raw data and comparison change, respectively) show pronounced changes in the fatigue, anticipation, and anxiety levels, with the most significant being a 60.4% increase in anxiety levels, a 53.8% increase in anticipation, and a 49.4% increase in tiredness.
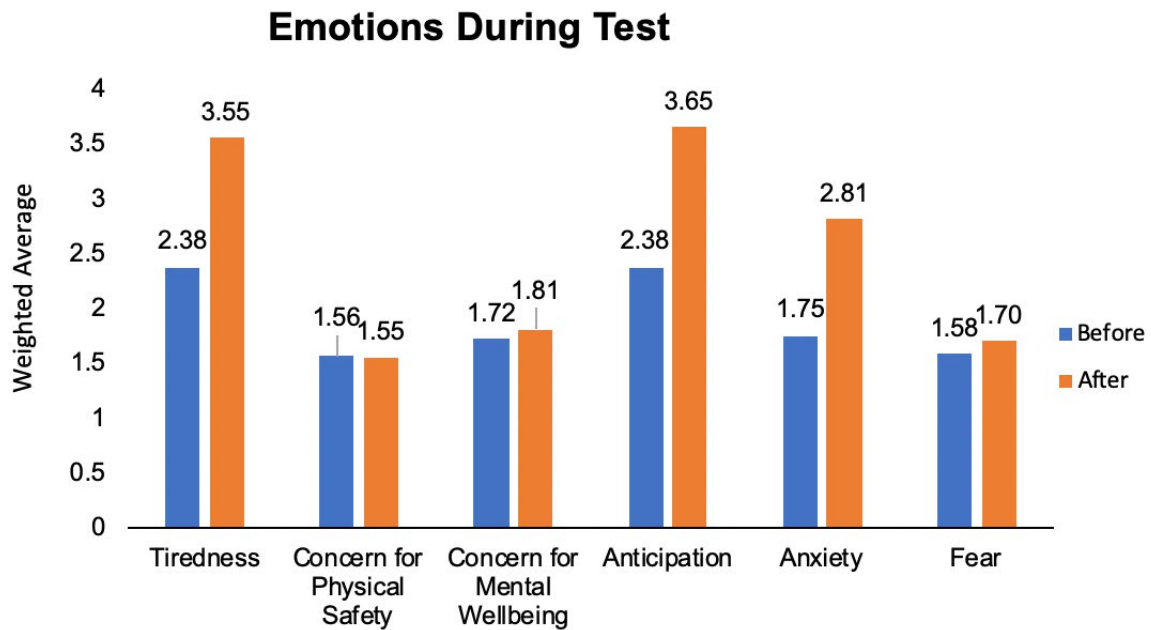
Emotions During Test

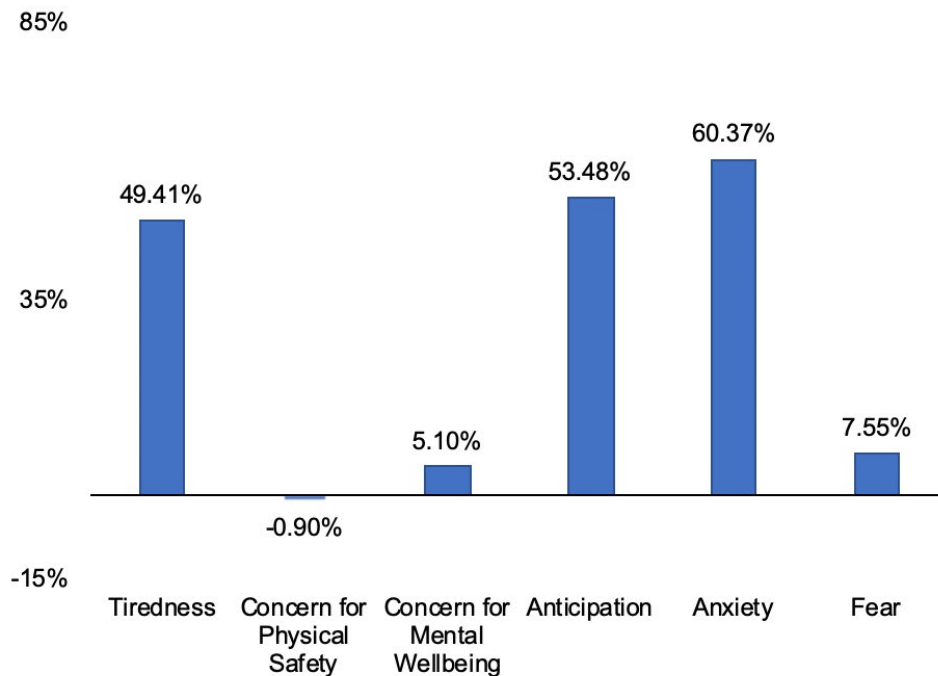**Figure 43. Reported Emotional States Before and After the Experiment**

Figure 44. Changes in Emotional State During the Test

With respect to physical effects, the researchers asked participants to rate how nauseous they felt during the simulation and whether nausea affected their ability to take control of the vehicle following disengagement in highway scenarios. The authors did not find any statistical significance in performance between the participants who experienced nausea and those who didn't. A majority (87.1%) of the participants felt some level of nausea during the test, ranging from mild to moderate. Figure 45 summarizes the findings and shows that

40% of the participants expressed that nausea affected their driving, while 50% of those who declared experiencing nausea mentioned it did not affect their driving at all. Out of those who claimed that nausea affected their driving, 25% did not successfully recover from the disengagement—a number significantly lower than in the previous study (81%). The average drift of all the participants who expressed that nausea affected their driving was 1.189 m, which was only slightly higher than the average maximum drift across all tests (1.193 m). Nausea is known to have the potential to create bias in the results of simulated studies, and this is one of the reasons why the researchers opted to set up the novel metric of the integral offset ratio: this way, each participant would have their own simulated baseline (possibly affected by queasiness) to which to compare their performance.

Figure 45. Nausea Effect Reported During Self-Assessment

V. SUMMARY AND CONCLUSIONS

Personal devices are a widespread commodity that is increasing its presence in vehicles with active engagement on the part of the driver while at the wheel. Trends for driver distractions are worrisome, especially for young drivers. In 2017, NHTSA estimated over 21,000 deaths on the road related to cell-phone distraction. No national ban on cell-phone usage while at the wheel has been issued for the United States, and many state-level regulations allow the use of personal devices if operated hands-free and secured in the vehicle. The hands-free ban limits manual engagement, thus preventing a level of physical distraction; it does not prevent, however, visual and cognitive distraction from the driving task. Solid scientific evidence to quantify the safety improvement or distraction reduction brought along by hands-free interactions and voice-to-text capabilities is not yet consolidated within the literature. This study thus aimed at investigating how different modes of interaction of drivers with a smart phone (i.e., manual texting versus vocal input) affect drivers' distraction and performance in both conventional and semi-autonomous vehicles.

Different variables were measured for two scenarios, which included a suburban scenario with a regulated intersection stop subject to the reception of two text messages, and a highway scenario with autonomy disengagement at a highway exit subject to the reception of one text message. The engineering variables measured and investigated are summarized in Table 16.

Table 16. Summary of Investigated Variables for the Study

Scenario	Time-to-Input Variables	Vehicle Trajectory Variables	Eye Tracking Variables
Suburban intersections and traffic light	Time-to-throttle: response time to the stimulus of the traffic light turning green in terms of acceleration input	Stopping distance behind a vehicle stopped at the red traffic light	▪ Time spent looking at the phone vs. at the road ▪ Maximum dwelling time of gaze on cell phone
Highway exit during automation disengagement	• Time-to-throttle • Time-to-steer • Time-to-brake (all computed from the disengagement point)	• Lateral offset from the center of the lane during takeover • Integral ratio to compare conventional driving to driving after disengagement	▪ Time spent looking at the phone vs. at the road ▪ Maximum dwelling time of gaze on cell phone

Two-tailed t-tests were used to analyze statistical difference for all the variables investigated as a function of the two tested interfaces (i.e., manual text reply and vocal reply). Table 17 provides a summary of the statistical significance or non-significance for the investigated variables.

Table 17. Summary of Statistical Significance for the Investigated Variables

Suburban Scenario		Highway Scenario	
Variables found to be not statistically significant	Variables found to be statistically significant	Variables found to be not statistically significant	Variables found to be statistically significant
Time-to-throttle	Time to interact (marginal)	Time-to-input (steering, throttle, braking)	Time to compose
Stopped distance from traffic	Time to compose	Drift and integral ratio	ETA recollection (marginal)
Speed at which the phone was first engaged with	ETA recollection (marginal)	Dwelling time on phone	Perceived effort for texting
Dwelling time on phone	Perceived mental workload for texting (marginal)	Dwelling time on windscreen	
Dwelling time on windscreen	Perceived effort for texting	Percentage time on phone	
Percentage time on phone		Percentage time on windscreen	
Percentage time on windscreen		Speed recollection	
Speed recollection			

The main conclusion of the study rests with the different nature of those variables that were found to be statistically significant and those that were not. In particular, two types of variables were found to be statistically significant:

1. **Texting-related variables:** composition time for a text was found to be lower for the vocal interface. Length of text in terms of characters contained was similar for both interfaces, but the vocal interface (as intuitively predicted) gave a shorter time of text composition. In other words, people were faster at composing a reply through the vocal interface. Furthermore, for the suburban scenario, marginal significance was found for the time of first click, or time-to-interaction, with people more likely to tap on the phone sooner with the manual interface.

2. **Subjective measures:** these are variables that are self-assessed by the participants through the surveys administered after the tests. In particular, lower perceived effort was found for the vocal interface, and participants expressed a clear preference for the vocal interface, indicating that the level of compromised safety was higher for the manual interface.

Despite the above, there was no statically significant improvement[34] seen in participants' performance with respect to the objective engineering metrics investigated in the study, including response times and drift/lateral offset. This conclusion has implications for the current state of driving regulations in the U.S., which overwhelmingly enforce a texting ban while making no stipulations regarding the use of voice-reply technology. The intuitiveness of a hand-held ban was confirmed by the study, as reported in finding (2) above. It was interesting to note that the participants confirmed both before and after the test that they had confidence in the fact that the vocal interface was more intuitive, easier to use, and would imply a lower risk and lower safety compromise. The authors believe that this personal belief traces along the same line of thought that intuitively corroborates the usefulness of

having the hand-held device ban, which 21 U.S. states currently adopt. At the same time, it was even more interesting to observe the lack of statistically significant difference in the measured quantitates (speed of responses, vehicle trajectory, and eye tracking). This lack indicates that the authors could not prove a safety improvement when voice-to-text interfaces are employed.

The authors thus bring forward the careful recommendation to further study hands-free interfaces and be wary of the technological promises of vocal engagement interfaces coming out for infotainment purposes. Nevertheless, the results of this report are to be considered preliminary and limited in scope. Many factors hamper the generalizability of this study, including:

- The small sample size: 32 participants completed the tests, for a total of 128 data points. Statistically speaking, the test is considered sound considering that only one independent variable with two levels was considered (i.e., interface, with levels being manual or vocal). However, having opted for a within-subject design where every participant got exposed to the four possible combinations of scenarios and interface, learning has to be taken into account. The authors opted to employ randomization of the testing scenarios to limit those effects, but a larger sample size could have displayed different behaviors.

- The fidelity of the simulator: while the simulator employed in the study is NHTSA-approved and represent state-of-the-art technology, there is no guarantee that people exposed to similar scenarios in real-life and real-roads will behave in the exact same way. Eriksson and co-authors proved comparability of results in terms of response times after disengagements for simulator studies and real-roads studies,[35] but no current methodology exists to translate and/or create a correspondence between results obtained in a simulator and those expected in real-life.

- The particular HMI tested: different designs of the human interface could generate different results. Within our HMI, participants still had to tap on the phone to initiate any type of reply, regardless of manual or vocal entry of the text. All-vocal interfaces, that presume no need to ever tap on the screen, are being released to the public, and could (in the long-run) generate different modality of interactions with personal devices. A study spanning several designs of HMIs would be needed to assess and estimate how much results can be expected to change when different interface options are being analyzed.

APPENDIX A: IN-BETWEEN SURVEY

Post-Highway Survey

Post Highway Questionnaire

Participant ID: Examiner:
Randomization: ☐ High. m ☐ High. v   Date / Time:
Will be filled out by the examiner

Workload:

1. Mental Demand: How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)?

Low   High

2. Physical Demand: How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)?

Low   High

3. Temporal Demand: How much time pressure did you feel due to the rate of pace at which the task elements occurred?

Low   High

4. Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter?

Good   Poor

5. Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low   High

6. Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Low   High

Situation Awareness:

7. Did you see one of the following animals while you were driving?

○ Giraffe ○ Lion ○ Elephant ○ Horse ○ None of them

8. Do you remember the estimated time of arrival?

_____

9. How fast was the vehicle traveling at the time of the disengagement?

_____

10. Is your answer to the previous question based on an actual reading of the speedometer?

○ Yes ○ No

Disengagement:

11. What were your first instincts at the time of the disengagement?
(Check all boxes that apply)

☐ Accelerating   ☐ Brake
☐ Reach for the steering wheel   ☐ I do not remember
☐ Other: _____
_____

12. In which direction did you initially steer the vehicle after the take over request?

○ Left   ○ Straight
○ Right   ○ I do not remember
○ Other: _____
_____

13. Do you believe you were successful in recovering from the disengagement while remaining the lane of travel?

○ Yes ○ No ○ I do not remember

Location:

14. Which lane was the vehicle in during the disengagement?

○ Left lane ○ Middle lane ○ Right lane
○ I do not remember ○ No disengagement happened

Please turn the sheet!

15. How far from the center of the exit lane did the vehicle deviate (maximum deviation)? (% of lane width)?

○ No deviation (0-10 % lane width) ○ Minimal deviation (10-20%)
○ Moderate (20-40%) ○ Substantial, near the border of the lane (40-50%)
○ Loss of control (more than 50%) ○ I do not remember

16. Did you feel you maintained safety safe distances from traffic during this portion of the simulation?

○ Yes ○ No ○ I do not remember

Phone:

17. Did you finish replying to the text message before taking control of the vehicle?

○ Yes ○ No ○ I do not remember

18. Do you feel like you were prioritizing the responding to the text message or driving the vehicle?

○ prioritized the responding to the text ○ prioritized driving the vehicle ○ I do not remember

19. Do you think responding to the text while driving compromised safety?

     1   2   3   4   5
Not at all ○ ○ ○ ○ ○ Yes, of course

20. Did you notice the incoming text or the disengagement first?

○ incoming text ○ light turning red ○ I do not remember

Eye Gaze:

21. What were you looking at when you received the text message?

○ Phone/Text ○ Phone/GPS ○ Dashboard
○ Mirror ○ Road ○ Other vehicles
○ Center console ○ Steering wheel
○ Other: _____

22. Where was the first place you looked just immediately after receiving the text?

_____
_____
_____

23. What were you looking at when the take-over request was issued?

○ Phone/Text ○ Phone/GPS ○ Dashboard
○ Mirror ○ Road ○ Other vehicles
○ Center console ○ Steering wheel
○ Other: _____

24. While replying to the text message, which strategy did you adopt?

○ Short alternating gazes between phone and outside traffic
○ Focus on the phone first and on outside environment after
○ Focus on the outside environment first and the phone after
○ Sustained (long) focus time on the outside environment with short peeks at the phone
○ Other: _____
_____

25. Please rate your gaze fixation ratio on the phone compared to the environment?

	100% on the phone	75% on phone & 25% on the environment	50% on the phone & 50% on the environment	25% on the phone & 75% on the environment	100% on the environment
Ratio	○	○	○	○	○

Thank You!

Post-City/-Suburban Survey

Post City Questionnaire

Participant ID: _____ Examiner: _____

Randomization: ☐ City m ☐ City v Date / Time: _____

Will be filled out by the examiner

Workload:

1. Mental Demand: How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)?

Low — High

2. Physical Demand: How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)?

Low — High

3. Temporal Demand: How much time pressure did you feel due to the rate of pace at which the task elements occurred?

Low — High

4. Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter?

Good — Poor

5. Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low — High

6. Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Low — High

Situation Awareness:

7. Did you see one of the following animals while you were driving?

○ Giraffe ○ Lion ○ Elephant ○ Horse ○ None of them

8. Do you remember the estimated time of arrival?

_____

9. What do you believe was your average traveling speed?

_____

10. Is your answer to the previous question based on an actual reading of the speedometer?

○ Yes ○ No

Location:

11. Did you successfully stop at the traffic light (respected safety distances)?

○ Yes ○ No ○ I do not remember

12. Did you feel you maintained a safe distance between your vehicle and the stopped vehicle in front of you at the red light?

○ Yes ○ No ○ I do not remember

Phone:

13. Did you finish replying to the FIRST text message before coming to a stop at the light?

○ Yes ○ No ○ I do not remember

14. Do you feel like you were prioritizing the responding to the text messages or driving the vehicle?

○ prioritized the responding to the text ○ prioritized driving the vehicle ○ I do not remember

15. Where was the first place you looked just immediately after receiving the FIRST text?

_____
_____
_____

16. Did you finish replying to the SECOND text before resuming to drive (after the light turned green)?

○ Yes ○ No ○ I do not remember

Please turn the sheet!

17. Do you think responding to the text while driving compromised safety?

Not at all 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ Yes, of course

18. Did you notice the first incoming text or the light turning red first?

○ first incoming text ○ light turning red ○ I do not remember

19. Did you notice the second incoming text or the light turning green first?

○ second incoming text ○ light turning green ○ I do not remember

Eye Gaze:

20. What were you looking at when you received the FIRST text message?

○ Phone/Text ○ Phone/GPS ○ Dashboard
○ Mirror ○ Road ○ Other vehicles
○ Center console ○ Steering wheel ○ Traffic light
○ Other: _____

21. While replying to the FIRST text message, which strategy did you adopt?

○ Short alternating gazes between phone and outside traffic
○ Focus on the phone first and on outside environment after
○ Focus on the outside environment first and the phone after
○ Sustained (long) focus time on the outside environment with short peeks at the phone
○ Other: _____

22. Please rate your gaze fixation ratio on the phone compared to the environment?

	100% on the phone	75% on phone & 25% on the environment	50% on the phone & 50% on the environment	25% on the phone & 75% on the environment	100% on the environment
Ratio	○	○	○	○	○

Thank You!

ABBREVIATIONS AND ACRONYMS

CM	City Manual
CV	City Vocal
ETA	Estimated Time of Arrival
FN	False Negative
FP	False Positive
HITL	Human in the Loop
HM	Highway Manual
HMI	Human Machine Interface
HV	Highway Vocal
IRB	Institutional Review Board
M	Mean
MTI	Mineta Transportation Institute
NASA	National Aeronautics and Space Agency
NHTSA	National Highway Traffic Safety Administration
PDF	Probability Density Function
SD	Standard Deviation
SJSU	San José State University
TLX	Task Load Index
TN	True Negative
TOR	Takeover Request
TP	True Positive
US	United States

ENDNOTES

1. "Federal Automated Vehicle Policy – Accelerating the Next Revolution in Roadway Safety," U.S. Department of Transportation, September 2016; NHTSA, National Highway Traffic Safety Administration: NHTSA – Distracted Driving 2015 – DOT HS 812 381, https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812381.

2. Zendrive, 2018 Distracted driving snapshot, http://blog.zendrive.com/distracted-driving-is-far-worse/; NHTSA data available at https://www.nhtsa.gov/risky-driving/distracted-driving.

3. "The Dangers of Distracted Driving," U.S. Federal Communications Commissions, https://www.fcc.gov/consumers/guides/dangers-texting-while-driving.

4. "Distracted Driving, Cell Phone Use and Texting While Driving Laws Maps," National Conference of State Legislature, http://www.ncsl.org/research/transportation/spotlight-distracted-driving.aspx.

5. "Distracted Driving, Cell Phone Use and Texting While Driving Laws Maps," National Conference of State Legislature, http://www.ncsl.org/research/transportation/spotlight-distracted-driving.aspx.

6. J. K. Caird, et al., "A meta-analysis of the effects of cell phones on driver performance," *Accident Analysis & Prevention* 40 (2008): 1282–1293.

7. "Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian - Vehicle Operator Role," National Transportation Safety Board, https://www.ntsb.gov/news/events/Documents/2019-HWY18MH010-BMG-presentation2.pdf.

8. SAE, Standard, S.A.E., J3016 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, Standard, SAE, USA.

9. Apple CarPlay, the ultimate Co-Pilot, July 2018, https://www.apple.com/ios/carplay/.

10. Drive Mode, July 2018, https://drivemode.com/.

11. P. R. Bakhit, Guo, B. and Ishak, S., 2018, "Crash and Near-Crash Risk Assessment of Distracted Driving and Engagement in Secondary Tasks: A Naturalistic Driving Study," *Transportation Research Record*, p.0361198118772703; "NHTSA–National Telephone Survey on Distracted Driving Attitudes and Behaviors 2015," National Highway Traffic Safety Administration, DOT HS 812 437, https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/traffic_tech/13118-2015_handheld_survey_031418_v4-tag.pdf; G. M. Fitch, et al., 2013, "The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk" (No. DOT HS 811 757).

12. G. M. Fitch, et al., 2013, "The impact of hand-held and hands-free cell phone use on

driving performance and safety-critical event risk" (No. DOT HS 811 757).

13. C. Neubauer, Matthews, G. and Saxby, D., 2012, September, "The effects of cell phone use and automation on driver performance and subjective state in simulated driving," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 56, No. 1), Sage CA: Los Angeles, CA: Sage Publications, 1987–1991.

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

15. G. M. Fitch, et al., 2013. The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk (No. DOT HS 811 757).

16. P. R. Bakhit, Guo, B. and Ishak, S., 2018. Crash and Near-Crash Risk Assessment of Distracted Driving and Engagement in Secondary Tasks: A Naturalistic Driving Study, *Transportation Research Record*, p.0361198118772703.

17. R. E. Llaneras, Salinger, J. and Green, C.A., 2013. Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway. In: Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 17–20, 2013, Bolton Landing, New York. Iowa City, IA: Public Policy Center, University of Iowa, 2013: 92–98.

18. T. M. Pickrell, & Ye, T. J. (2011, December). Driver Electronic Device Use in 2010. (Traffic Safety Facts Research Note. Report No. DOT HS 811 517). Washington, D.C.: National Highway Traffic Safety Administration, www-nrd.nhtsa.dot.gov/Pubs/811517.pdf.

19. AAA, American Automobile Association, "Effectiveness of Distracted Driving Countermeasures: A Literature Review," November 2019, https://aaafoundation.org/wp-content/uploads/2019/11/19-0553_AAAFTS-DD-Countermeasures-Brief_r3.pdf.

20. "Governors Highway Safety Association - Distracted Driving," available at https://www.ghsa.org/index.php/state-laws/issues/distracted%20driving.

21. Young, Kristie, Michael Regan, and M. Hammer. "Driver distraction: A review of the literature." *Distracted driving* 2007 (2007): 379-405.

22. Barón, Adriana, and Paul Green. *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review*. No. UMTRI-2006-5. Ann Arbor, MI: University of Michigan, Transportation Research Institute, 2006.

23. Jibo He, et al., "Texting while driving: Is speech-based text entry less risky than handheld text entry?" *Accident Analysis & Prevention* 72 (2014): 287–295.

24. Sandra G. Hart, and Lowell E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in Psychology* 52: 139–183. North-Holland, 1988.

25. Sandra G. Hart, and Lowell E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in Psychology* 52: 139–183. North-Holland, 1988.

26. R. E. Llaneras, Salinger, J. and Green, C.A., 2013. Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway. In Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 17–20, 2013, Bolton Landing, New York. Iowa City, IA: Public Policy Center, University of Iowa, 2013: 92–98.

27. U.S. Department of Transportation – Federal Highway Administration FHWA, 2016. Licensed Drivers in the US Demographics for 2016, available at https://www.fhwa.dot.gov/ohim/onh00/bar7.htm.

28. 28　　　Saccades are rapid jittery movements of both eyes from one point of fixation to another

29. Katja Kircher, Albert Kircher, and Fredrich Claezon. "Distraction and drowsiness-a field study." VTI, Swedish National Road and Transport Research Institute, Linköping, Sweden (2009).

30. W. Zhu et al. (2010), "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations," NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland, 19, 67.

31. Francesca M. Favarò, et al., "Analysis of Disengagements in Semi-Autonomous Vehicles: Drivers' Takeover Performance and Operational Implications" (MTI, 2019).

32. B. Mok et al. 2015, "Emergency, Automation Off: Unstructured Transition Timing for Distracted Drivers of Automated Vehicles," in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2458–64, https://doi.org/10.1109/ITSC.2015.396.

33. Francesca M. Favarò, et al. "Analysis of Disengagements in Semi-Autonomous Vehicles: Drivers' Takeover Performance and Operational Implications" (MTI, 2019).

34. This means that a statistical test that analyzes two distributions cannot ascertain that they are not two instances of the same original distribution. In other words: through the experiment, the authors obtained data from the manual tests and the vocal tests, in the form of two distributions. When comparing them, the statistical tests failed to verify that the two distributions are different. This is a common practice to analyze results from any type of user-study or survey.

35. Eriksson, Alexander, V. A. Banks, and N. A. Stanton. "Transition to manual: Comparing simulator with on-road control transitions." *Accident Analysis & Prevention* 102 (2017): 227-234.

BIBLIOGRAPHY

AAA, American Automobile Association. Effectiveness of Distracted Driving Countermeasures: A Literature Review. November 2019. Available at https://aaafoundation.org/wp-content/uploads/2019/11/19-0553_AAAFTS-DD-Countermeasures-Brief_r3.pdf.

Apple CarPlay, the ultimate Co-Pilot. July 2018. https://www.apple.com/ios/carplay/.

Bakhit, Peter R., BeiBei Guo, and Sherif Ishak. "Crash and Near-Crash Risk Assessment of Distracted Driving and Engagement in Secondary Tasks: A Naturalistic Driving Study." *Transportation Research Record: Journal of the Transportation Research Board* 2672 (2018): 245–54. https://doi.org/10.1177/0361198118772703.

Barón, Adriana, and Paul Green. *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review*. No. UMTRI-2006-5. Ann Arbor, MI: University of Michigan, Transportation Research Institute, 2006.

Caird, Jeff K., Kate A. Johnston, Chelsea R. Willness, Mark Asbridge, and Piers Steel. "A Meta-Analysis of the Effects of Texting on Driving." *Accident Analysis & Prevention* 71 (2014): 311–18. https://doi.org/10.1016/j.aap.2014.06.005.

Drive Mode. July 2018. https://drivemode.com/.

Eriksson, A., V. A. Banks, and N. A. Stanton. "Transition to manual: Comparing simulator with on-road control transitions." *Accident Analysis & Prevention* 102 (2017): 227-234.

Fitch, G. A., S. A. Soccolich, F. Guo, J. McClafferty, Y. Fang, R. L. Olson, M. A. Perez, R. J. Hanowski, J. M. Hankey, & T. A. Dingus. 2013. "The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk." (Report No. DOT HS 811 757). Washington, D.C.: National Highway Traffic Safety Administration.

Governors Highway Safety Association. 2000. "Distracted Driving." Ghsa.Org. GHSA. 2000. Available at https://www.ghsa.org/state-laws/issues/distracted%20driving.

Hart, Sandra G., and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." In *Advances in Psychology* 52: 139–183. North-Holland, 1988.

Kircher, Katja, Albert Kircher, and Fredrich Claezon. "Distraction and drowsiness-a field study." VTI, Swedish National Road and Transport Research Institute, Linköping, Sweden (2009).

Llaneras, R.E., Salinger, J. and Green, C.A., 2013. Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention

to the forward roadway. In Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 17–20, 2013, Bolton Landing, New York. Iowa City, IA: Public Policy Center, University of Iowa: 92–98.

Mok, B, Johns, Lee, Miller, Sirkin, Ive, and Ju. 2015. "Emergency, Automation Off: Unstructured Transition Timing for Distracted Drivers of Automated Vehicles." In 2015 IEEE 18th International Conference on Intelligent Transportation Systems: 2458–64. IEEE. https://doi.org/10.1109/ITSC.2015.396.

National Conference of State Legislature. Distracted Driving, Cell Phone Use and Texting While Driving Laws Maps. Available at http://www.ncsl.org/research/transportation/spotlight-distracted-driving.aspx

National Highway Traffic Safety Administration, NHTSA. Distracted Driving 2015 – DOT HS 812 381. Available online at https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812381

National Highway Traffic Safety Administration, NHTSA. National Telephone Survey on Distracted Driving Attitudes and Behaviors 2015– DOT HS 812 437. Available online at https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/traffic_tech/13118-2015_handheld_survey_031418_v4-tag.pdf.

National Transportation Safety Board. Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian - Vehicle Operator Role. Available at https://www.ntsb.gov/news/events/Documents/2019-HWY18MH010-BMG-presentation2.pdf.

Neubauer, C., Matthews, G. and Saxby, D. 2012, September. The effects of cell phone use and automation on driver performance and subjective state in simulated driving. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting 56: 1987–1991. Sage CA: Los Angeles, CA: Sage Publications.

Pickrell, T. M., and T. J. Ye (2011, December). Driver Electronic Device Use in 2010. (Traffic Safety Facts Research Note. Report No. DOT HS 811 517). Washington, D.C.: National Highway Traffic Safety Administration. Available at www-nrd.nhtsa.dot.gov/Pubs/811517.pdf.

SAE, Standard, S.A.E. J3016 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. Standard. SAE, USA.

U.S. Department of Transportation, September 2016. Federal Automated Vehicle Policy: Accelerating the Next Revolution in Roadway Safety. NHTSA, National Highway Traffic Safety Administration.

U.S. Department of Transportation, Federal Highway Administration FHWA. 2016. Licensed Drivers in the US Demographics for 2016. Available at https://www.fhwa.

dot.gov/ohim/onh00/bar7.htm

U.S. Federal Communications Commissions. The Dangers of Distracted Driving. Available at https://www.fcc.gov/consumers/guides/dangers-texting-while-driving

Young, K., Regan, M. and M. Hammer. "Driver distraction: A review of the literature." *Distracted driving* 2007 (2007): 379-405

Zendrive. 2018 Distracted driving snapshot. Available at http://blog.zendrive.com/ distracted-driving-is-far-worse/ and NHTSA data used available at https://www. nhtsa.gov/risky-driving/distracted-driving

Zhu, W., N. Zeng, and N. Wang. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland 19, 67.

ABOUT THE AUTHORS

FRANCESCA M. FAVARÒ

Dr. Favarò is an Assistant Professor in the Department of Aviation and Technology in the College of Engineering at San José State University. Prior to joining SJSU, she earned a Ph.D. and M.S. in Aerospace Engineering at the Georgia Institute of Technology and an M.S. and B.S. in Space Engineering at Politecnico di Milano, Italy. Dr. Favarò's research interests lie in the broad field of system safety and risk analysis, with an emphasis on system engineering concepts and the safe integration and embedding of new technologies and the consistent update of regulations and certification practices. In 2016, she funded the RiSA^{2}S lab, which deals with Risk and Safety Assessment of Autonomous Systems such as unmanned aircraft systems and self-driving cars. In 2017 she became an MTI research associate and started collaborating as an expert in the realm of autonomous vehicles. Her interests are currently focused on the safe integration of autonomous systems within U.S. public roads as well as national airspace. She particularly focuses on bridging the gap between the technological world and the current regulatory panorama. Dr. Favarò has authored several journal publications and conference proceedings on a variety of topics, ranging from UAS mishaps to risk assessment and regulatory compliance for nuclear power plants. Dr. Favarò serves on the Board of Directors of the Aero Club of Northern California and acts as Secretary for CAAPSO (Community and Airport Partnership for Safe Operations), an aviation interest group that safeguards safe and efficient operations at the local airport of Reid-Hillview, San José. She is an FAA-certified Advanced Ground Instructor, a certified Remote Pilot for UAS commercial operations, and a solo-endorsed pilot.

PEER REVIEW

San José State University, of the California State University system, and the Mineta Transportation Institute (MTI) Board of Trustees have agreed upon a peer review process required for all research published by MTI. The purpose of the review process is to ensure that the results presented are based upon a professionally acceptable research protocol.