SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Analysis of Disengagements in Semi-Autonomous Vehicles: Drivers' Takeover Performance and Operational Implications

Francesca M. Favaro, PhD
Sky Eurich
Syeda Rizvi
Shivangi Agarwal
Sumaid Mahmood
Nazanin Nader

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 PhD, 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 19-15

ANALYSIS OF DISENGAGEMENTS IN SEMI-AUTONOMOUS VEHICLES: DRIVERS' TAKEOVER PERFORMANCE AND OPERATIONAL IMPLICATIONS

Francesca M. Favaro, PhD
Sky Eurich
Syeda Rizvi
Shivangi Agarwal
Sumaid Mahmood
Nazanin Nader

June 2019

A publication of
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 19-15	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Analysis of Disengagements in Semi-Autonomous Vehicles: Drivers' Takeover Performance and Operational Implications		**5. Report Date** June 2019 **6. Performing Organization Code**
7. Authors Favaro, Francesca https://orcid.org/0000-0003-0239-0770 Eurich, Sky https://orcid.org/0000-0001-6984-4875 Rizvi, Syeda https://orcid.org/0000-0002-5655-1835 Mahmood, Sumaid https://orcid.org/0000-0003-1127-8836 Nader, Nazanin https://orcid.org/0000-0002-1159-4591		**8. Performing Organization Report** CA-MTI-1710
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

16. Abstract

This report analyzes the reactions of human drivers placed in simulated Autonomous Technology disengagement scenarios. The study was executed in a human-in-the-loop setting, within a high-fidelity integrated car simulator capable of handling both manual and autonomous driving. A population of 40 individuals was tested, with metrics for control takeover quantification given by: i) response times (considering inputs of steering, throttle, and braking); ii) vehicle drift from the lane centerline after takeover as well as overall (integral) drift over an S-turn curve compared to a baseline obtained in manual driving; and iii) accuracy metrics to quantify human factors associated with the simulation experiment. Independent variables considered for the study were the age of the driver, the speed at the time of disengagement, and the time at which the disengagement occurred (i.e., how long automation was engaged for). The study shows that changes in the vehicle speed significantly affect all the variables investigated, pointing to the importance of setting up thresholds for maximum operational speed of vehicles driven in autonomous mode when the human driver serves as back-up. The results shows that the establishment of an operational threshold could reduce the maximum drift and lead to better control during takeover, perhaps warranting a lower speed limit than conventional vehicles. With regards to the age variable, neither the response times analysis nor the drift analysis provide support for any claim to limit the age of drivers of semi-autonomous vehicles.

17. Key Words Autonomous vehicles; transportation safety; driver performance; reaction time; drift	**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)

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu

063019

ACKNOWLEDGMENTS

Funding for this research was provided by the US Department of Transportation Office of the Assistant Secretary for Research and Technology University Transportation Centers Program (grant 69A3551747127). The grant is managed by the Mineta Transportation Institute (MTI) of San Jose, CA. The authors would like to thank MTI Executive Director, Dr. Karen Philbrick, as well as Dr. Hilary Nixon, Deputy Executive Director, for all the help and support during the proposal submission and study. The authors would also like to thank Research Support Assistant Joseph Mercado, Executive Administrative Assistant Jill Carter, and Editing Press for editorial services.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Autonomous Vehicle (AV) technology is quickly expanding its market. Manufacturers are targeting different levels of autonomy, with semi-autonomous vehicles currently in the lead. In semi-autonomous vehicles, a human driver collaborates with the software that acts as "brain" of the vehicle and serves as back-up whenever the Autonomous Technology (AT) disengages after a failure. Current regulations require the human driver to monitor the safe operation of the vehicle at all times, and to be capable of taking over immediate control in the event of an autonomous technology failure. In the safety-critical situation of an AT disengagement, it is important to ensure that the human driver has enough time to react and respond effectively to the vehicle request for human control.

This study analyzed the reactions of human drivers placed in simulated AT disengagement scenarios. The study was executed in a human-in-the-loop setting, within a high-fidelity integrated car simulator capable of handling both manual and autonomous driving. A population of 40 individuals was tested for control takeover metrics quantified as: response times (considering inputs of steering, throttle, and braking); vehicle drift from the lane centerline after takeover as well as integral drift over an S-turn curve compared to a baseline obtained in manual driving; and accuracy metrics to quantify recollection and situational awareness.

Independent variables considered for the study were the age of the driver, the speed at the time of disengagement, and the time at which the disengagement occurred (i.e., how long automation is engaged for). These three independent variables were chosen in order to answer specific operational questions in relation to the use of semi-AVs on US public roads: Will there be constraints on the maximum speed at which the systems can be safely operated? Will there be constraints on the maximum duration for which the system can be operated safely? Should semi-AVs be sold only to people in a certain age group (not too young/not too old)?

Drivers completed the tests in a simulated 7.6 mile closed-loop track that resembled a highway environment. The observations collected from the 40 tests have to be considered preliminary in nature given the small sample size, but nonetheless show interesting results with important operational implications. Among the notable statistically significant results are the following:

1. Of the two speed settings selected for the study (high speed of 65 mph and low speed of 55 mph), the low-speed category yielded better performance for all test subjects. The average maximum drift in automated mode after disengagement increased from 3 ft at low speed to 6.5 ft at high speed. Additionally, variation between fully manual driving performance, and driving performance during manual takeover after AT disengagement, was smaller at the low-speed setting than at the high-speed setting. Higher speeds also led to more pronounced changes in the level of trust in the technology as well as higher reported nervousness and fear in the experience.

2. Success of the takeover maneuver was measured as a function of lateral drift. Participants were required to take over control of the vehicle after the disengagement,

and remain within the same lane of travel. In 69% of the cases, unintentional lane departures were recorded. Moreover, all participants but one still described their control takeover as “successful” (i.e., within the lane boundaries). Accuracy of the estimation of success in remaining within the lane was lower than 50%.

3. The duration of engagement of automation did not exhibit a linear trend for performance decrease. This means that the study did not show that a longer engagement led to a worsening performance. The selected dependent variables did not show statistically significant trends. Additional tests are needed to further investigate the dependency on duration of engagement.

4. Of the three age groups tested (18–35; 35–55; 55+), the age group of 55+ performed best in terms of both maximum drift and comparison between conventional driving and driving after AT failure.

5. An analysis of the first input used showed that 78% of the participants resorted to acceleration and steering rather than braking and decreasing the vehicle’s speed after control takeover following the disengagement. Recollection of the first input was also tested after the end of the simulation: 32% of those that first resorted to acceleration and steering incorrectly recollected braking to be their response to the disengagement event (i.e., they thought they had braked but they had accelerated instead).

6. Although all participants received both an auditory and a visual warning for the disengagement, 50% of them reported not seeing the visual icon, displayed in the 10.2-inch central console, that indicated to take back control of the vehicle. Indeed, 76% of the participants expressed a preference for HeadsUp Displays, which are just now making their way on the market.

7. Finally, we observed low accuracy in recollecting and estimating the speed of the vehicle, as well as a tendency to overestimate the duration of engagement of the technology.

From a regulatory standpoint, the preliminary results point to the importance of setting up thresholds for maximum operational speed of vehicles driven in autonomous mode when the human driver serves as back-up, perhaps warranting a lower speed limit than conventional vehicles. This research shows that the establishment of an operational threshold could reduce the maximum drift and lead to better control during takeover. Unintentional drift also attests to the need for discussions on possible dedicated lane usage for autonomous vehicles and separation from conventional traffic, as well as for the possibility of increasing lane width in dedicated lanes for semi-autonomous vehicles. With regards to the age variable, neither the response times analysis nor the drift analysis provide support for any claim to limit the age of drivers of semi-autonomous vehicles.

Wherever possible, the results were traced back to notable literature on the topic and were found to be in accordance. Future work will include further investigations of time of engagement, as well as validation of the results for bigger populations.

I. INTRODUCTION

BACKGROUND

Autonomous Vehicle (AV) technology is quickly expanding its market. Many factors account for the interest in this technology, including:

1. the improvement of the commute experience: self-driving transportation allows commuters to better allocate their commute time, and self-driving vehicles have the potential to shorten the commute once the car is able to take care of parking for itself, after the passenger has exited (Anderson et al., 2014);

2. the long-sought improvement of mobility for everyone, enabling differently abled people to access transportation and improving independence (US DoT, 2016);

3. the potential for fuel savings through optimized usage of braking and throttle, as well as more manageable parking arrangements, which help classify this type of technology as a “green” and eco-friendly alternative to more traditional means of transportation (Anderson et al., 2014); and

4. the potential safety improvement: recent statistics from the National Highway Traffic Safety Administration (NHTSA) attribute 94% of US crashes to human errors (Singh, 2015). Indeed, among the most frequently quoted advantages of AVs is the safety improvement that you might achieve once the “human element” is eliminated from the equation (Gao, Hensley & Zielke, 2014).

Manufacturers are targeting different levels of autonomy, with semi-autonomous vehicles currently in the lead. In semi-autonomous vehicles, a human driver collaborates with the software that acts as the “brain” of the vehicle and serves as back-up whenever the autonomous driving software (hereafter denoted as Autonomous Technology – AT) disengages after a failure. Figure 1 provides an overview of how the projected market for Autonomous Vehicles (AV), with estimated timelines and levels of automation targeted by several major manufacturers.

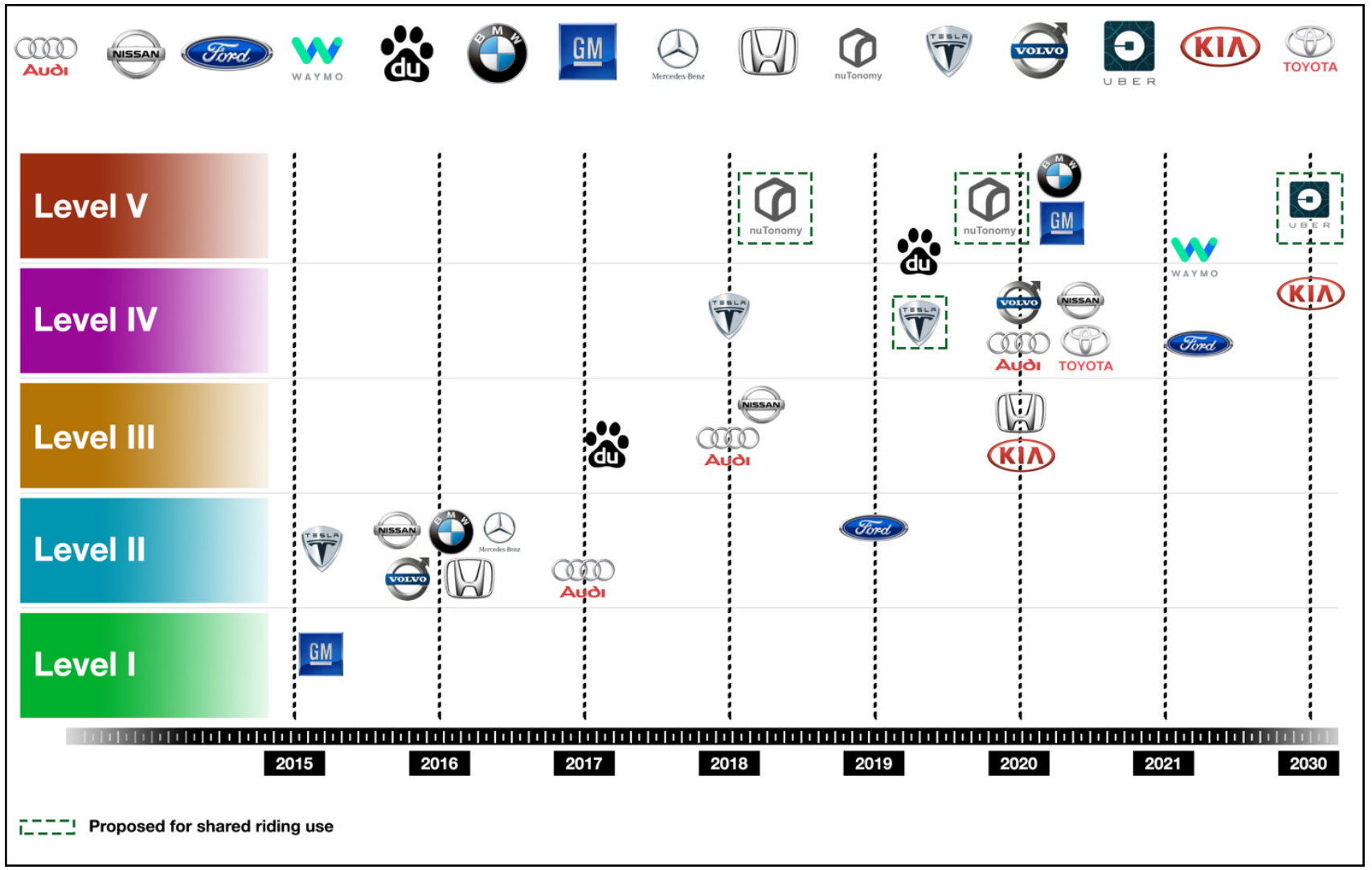

Figure 1. Overview of AV market, 2015–2030 Timeline (forecasted)

Note: Not meant to be exhaustive (Favaro et al., 2017).

The terms “semi-” and “fully-autonomous” are often informally used to distinguish between those AVs that require the presence of a human driver to operate and those that do not. This distinction, non-technical in nature, is based on the classification of levels of autonomy as defined by the Society of Automotive Engineers (SAE), and as reported in Figure 2, (SAE, 2014). SAE defined 6 levels of automation, ranging from Level 0 (no automation) to Level 5 (full unrestricted automation). The definition of the six levels (rows of Figure 2) is based on four factors (the four columns to the right of Figure 2) as follows:

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	**No Automation**	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	**Driver Assistance**	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	**Partial Automation**	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* (“system”) monitors the driving environment**						
3	**Conditional Automation**	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	**High Automation**	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	**Full Automation**	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

Figure 2. AV Levels of Automation

Note: Reproduced AS-IS with permission from SAE-International J3016TM (SAE, 2014).

1. The agent responsible for executing steering and throttle control: either human driver or AT;

2. the agent responsible for monitoring the external environment: either human driver or AT;

3. the agent responsible for serving as “back-up” when a failure prompts a disengagement of the AT: either human driver or AT; and

4. the driving modes in which autonomous operations are allowed: either “all modes of operations” unrestricted conditions or “some mode of operations” pre-specified conditions (e.g., good visibility).

Levels 1 through 3 are regarded as “semi-autonomous” due to the fallback back-up performance of the dynamic driving tasks placed on the human driver.

Whether forced by design choices or due to insufficient information regarding the context of a particular situation, an AV can enter into what it is called a “disengagement mode”. During disengagement, the full control and authority of the car movement is handed from the autonomous software to the human driver. Given that semi-AVs require collaboration between the AT and the human driver, the study of such interaction became of paramount importance in both the academic research world, the industry world, and the regulatory world.

Across the US, different states are in the process of creating ad hoc legislation for AVs (Fagnant & Kockelman, 2015; NCSL, 2017). In some states, fully autonomous technology, which does not require a driver at the steering wheel, is currently banned from public roads or limited to strict conditions of operation based on geo-fencing areas and weather conditions; for example, see the case of California, for both testing and deployment regulations (California Department of Motor Vehicles [CA DMV], 2017a; CA DMV, 2017b). The rationale behind such a choice stems from the low maturity of full-autonomous technology, and from opting for a conservative approach in order to limit the amount of technology deployed on public roads until clear guidelines have been established both at the Federal level as well as the state/local level.

As a result of the regulatory climate, which currently favors the gradual deployment of increasingly automated vehicles beginning with those that still require a human behind the steering wheel, a debate has started on the role of human drivers in the vehicles of the future, and on whether the presence of humans at the wheel may or may not be a “safer” choice than full-autonomy (Favaro et al. 2017; Davies, 2017).

SCOPE

One of the key aspects currently under examination by the research community is the interaction between the AT and the human driver. In particular, the study of such interaction is of paramount importance in all those situations that we may consider “off-nominal”, a term employed here to describe all those situations in which the authority of the vehicle switches from one agent to the other due to threats and hazards outside the regular operative conditions of the vehicle (e.g., a sudden request from the software for the human driver to regain control following a sensor malfunction, or following external conditions outside the AT capability—for instance presence of excessive pedestrians). In semi-autonomous vehicles, the human driver serves as back-up whenever the AT disengages following a failure. Current regulations place on the human driver the responsibility to carefully monitor the outside environment at all times (even when automation is engaged) and to be capable to immediately regain control should the vehicle request so (CA DMV, 2017a,b). At the same time, however, advanced autonomy allows the driver to not pay attention to the surroundings when the autonomous control is engaged (which is one of the main reasons these systems are advertised in the first place). It is then natural to wonder how safe it is to hand the control back to a potentially distracted human driver. In the safety-critical situation of an AT disengagement, it is important to ensure that the human driver has enough time to react and to respond effectively to the request to control the vehicle.

This report thus analyzes how human drivers placed in simulated AT disengagement scenarios respond to the emergency situation just described. The study was executed in a human-in-the-loop[1] setting, and examined drivers' responses to AT failures in semi-autonomous vehicles. A population of 40 individuals was tested, considering the following independent variables: age of the driver, speed at the time of disengagement, and time at which the disengagement occurred. These three independent variables were chosen to answer specific operational questions in relation to the use of semi-AVs on US public roads: Will there be constraints on the maximum speed at which the system can be safely operated? Will there be constraints on the maximum time for which the system can be continuously operated? Will there be constraints on the maximum time the system can be operated for? Should semi-AVs be sold only to people within a certain age range?

Participants received auditory and visual warning at the time of disengagement, and were asked to regain control of the vehicle, while maneuvering within a S-curve turn with instructions to remain within the original lane of travel. The study evaluated a number of dependent variables, including response times to the takeover request, drift performance, and several metrics to quantify human factors associated with recollection of the inputs used and situational awareness. These results are presented following a brief literature review on the topic, and a detailed discussion of the methodology employed for the study.

II. LITERATURE REVIEW

Table 1 presents a summary of notable studies that have been carried out in the past regarding the transfer of control authority from an AT device to a human driver. The studies included apply to AVs of comparable levels of autonomy to the level of autonomy used in this study. The table provides a summary of the available information regarding the response time/reaction time used, average computed values, information on external conditions, and overall study settings.

As can be gathered from the table, the majority of the studies were conducted in a simulator environment. Road testing is now also being conducted by AV manufacturers, and just recently in 2018, Waymo was the first manufacturer to obtain permission from the CA DMV to deploy its AV on public roads.

Table 1 also shows how heterogeneous the concept of “response time” can be, with some researchers opting to measure the first input provided by the driver, some only accounting for heading adjustments, and others actually measuring a reaction to a stimulus in the form of eye-gaze direction or hand movement. This is an important point that we will further address in our results section.

Table 1. A Summary of Notable Literature on the Topic of Control Handback and Driver Response

Reference	AV Level	Study Type (Road vs. Sim)	Definition of "reaction time" or "time to takeover" according to usage	Avg. Computed Time	Notes	Type of Disengagement	External Conditions
Bloomer, et al. 2017	Level 2 (ACC-adaptive cruise control, and LK-lane keeping assist)	Sim	Time until brake or steer onset, whichever came first	1.46 s	No distractor used. Reactions to forward collision avoidance in highly automated vehicle	Unstructured, sudden forward collision	Daytime driving in the right lane on a straight, 4 lane undivided highway in urban, rural and construction zone settings. Light traffic both in the oncoming and passing direction.
Shen & Neyens, 2017	Level 2	Sim	Time from initiated lane drift to first adjustment of heading	1.27 s no distractor 1.45 s with distractor	Distractor used: driver watching a movie	N/A	Gust of wind, lane departure
Forster, et al., 2017	Level 3	Sim	Reaction time: (1) time until button press (2) time until hand touches wheel (3) time until hands are available to use (4) gaze-reaction: first gaze fixation on road after takeover request	5.66 s –7.84 s hands on wheel 1.3s – 1.4 s gaze back on road	Distractor used: driver reading a magazine	Unstructured	No traffic, lane of the test vehicle splits into two lanes, causing the TOR (non-critical condition)
Payre, et al., 2016	Level 3 (fully automated driving)	Sim	First input by the participant (either braking or steering or gas)	4.3 s –8.7 s hands on wheel	No distractor used. Experimental campaign to test subsequent requests to takeover control of the vehicle	Structured during training phase (30 s warning), also structured during test drive, but the second time the vehicle disengaged the warning time was only 2 s	Four-lane highway (two in each direction), always straight section of road, disengagement was a sudden system failure

Reference	AV Level	Study Type (Road vs. Sim)	Definition of "reaction time" or "time to takeover" according to usage	Avg. Computed Time	Notes	Type of Disengagement	External Conditions
Zeeb., et al. 2016	Level 3 (conditionally automated driving)	Sim	(1) Time until eyes on road; (2) time until hands on steering wheel; (3) time until system deactivation (steering or braking)	1.5 s no distractor 1.4 s – 1.5 s with distractor	Outside visual cue provided. Distractors used: checking email, reading news, watching video	Structured, the automation doesn't immediately disengage, only after a four seconds following the **TOR (takeover request)**	Two-lane highway with traffic, disengagement was cause by either lane ending, construction or missing road markings
Petermeijer, et al., 2017	Level 3	Sim	(1) Time to touch steering wheel (2) Time to initiate steering wheel turn (3) Time until brake pedal was depressed (4) Steer touch or brake, minimum time of those two. Full response times: (5) Time to lane change (6) Time to car avoidance	1.6 s – 1.9 s hands on wheel and brake depression	Distractor used: visual search on tablet	Structured, TOR was given 7 seconds before an accident would occur.	Three-lane highway, cause of disengagement was a group of stationary cars in the road, required driver to steer around them, on straight road segments
Melcher, et al. 2015	Level 3	Sim	Time until first input, steering or braking	3.5 s	Distractor used: quiz game on mobile phone	Structured, System kept driving itself for 10 seconds after the TOR	TOR on highway, stopped vehicle in the lane, driver must steer around vehicle
Gold, et al. 2016	Level 3	Sim	Time until first input, steering or braking (different than time to hands on)	2.7 s –3.5 s	Different traffic conditions tested. Distractor used: verbal 20-question task	Unstructured	Varying levels of traffic

Reference	AV Level	Study Type (Road vs. Sim)	Definition of "reaction time" or "time to takeover" according to usage	Avg. Computed Time	Notes	Type of Disengagement	External Conditions
Dogan, et al. 2017	Level 2 (ACC and LK-traffic jam assist)	Sim	Takeover time is time until first input (brake or steer)	2.45 s	Different non-driving tasks considered	Unstructured	Sometimes vehicle would speed up past automation capabilities. Other times it would happen during low-speed traffic jams
Eriksson, et al. 2017	Level 2, Tesla S for road portion	Sim and Road	Time from onset of stimuli until action is complete (not until first input)	3.08 s on road 4.56 s sim	No distractor used	Unstructured	Simulated highway, and closed track for road portion
Zeeb, et al. 2015	Level 2 (ACC)	Sim	First user input, either braking or steering	2.09 s	Distractor used: texting or internet search	Unstructured, but visual cue, lane blocked by accident/ construction sign.	Two-lane highway, following behind SUV
Blanco et al., 2015	Level 2 and Level 3	Road	Time to first pre-defined input (steering or braking depending on experiment)	0.7 s –3.6 s	Distractor used: emails, GPS set-up, internet search	Structured (warning issued 30 s before actual disengagement)	Closed track
Mok, et al., 2015	Level 3	Sim	N/A	2.0 s –5.0 s*	*Minimum time needed between TOR and obstacle to overcome a lane closure	Unstructured but visual cue, lane closure	Closed track resembling highway environment

Note: Adapted from (Favaro et al., 2019).

Another important factor in Table 1 is the way the AT disengagement is set up within each study. When the need for a manual input arises, the AV generates a Take Over Request (TOR) to the human driver. TORs can have different forms, generally a combination of an aural and a visual warning (as it is in this study). Additionally, the TOR can be generated right at the time when the execution of manual input is needed, or as a prior warning (e.g., 30 seconds in advance) that the actual disengagement is about to happen soon.

Based on this distinction, we can speak of “structured disengagements”, where a prior-warning is issued to the driver before the disengagement takes place, and “unstructured disengagements”, where no prior warning is present. Historically, AT disengagements were first studied primarily in structured contexts, and the attention of the researching community has now shifted to unstructured ones. In this respect, however, note that although some studies claim to be executed in an “unstructured” setting, the presence of external visual threats serves a similar role to that of a fair warning. In other words, some of the Table 1 studies did not include a warning prior to the disengagements; however, visual cues such as a construction site or roads obstructions/lane closures were present in the simulation, possibly spiking the attentiveness of the driver and improving the response time compared to that of a purely unstructured and “unmotivated” disengagement (such as a sudden system failure).

The presence of external threats and visual cues was something that we avoided in this study. The setup used for our experimental campaign reflects a purely unstructured disengagement, with warnings to regain control provided to the driver only at the actual time of disengagement. Participants were also warned that a disengagement may or may not happen during their test.

Moreover, an important difference between the present work and previous literature is the set of independent variables investigated. The independent variables considered for this study were the age of the driver, the speed at the time of disengagement, and the time at which the disengagement occurred (i.e., how long automation was engaged for). There are no striking results in the literature with regard to any of these variables, and the ways in which they affect takeover performance after AT disengagement.

III. METHODOLOGY

This study employed human-in-the-loop (HITL) simulation with the aim of measuring the quality of control takeover of human drivers following a takeover request (TOR) issued by the system after a disengagement. To quantify such performance, the study employed a specific scenario of a TOR in a simulated highway environment, where participants sat in a vehicle driving in automated mode for a predetermined amount of time, after which a visual and auditory warning prompted the human driver to regain manual control of the vehicle. The importance of realistic HITL simulations stems from their ability to reproduce human errors. Although much of the functionality of automobiles can be automated, current and near-future semi-autonomous vehicles will still require human input into the system, meaning that there will still be the possibility of human-induced error (Treat et al., 1979; Favaro et al., 2017; Favaro et al. 2018). Understanding the limits of human capabilities in terms of monitoring and controlling semi-autonomous vehicles is an essential step in safely deploying these systems. To that end, this study aimed to use a high fidelity driving simulator to accurately assess the average driver's ability to recover from autonomous vehicle disengagements. This section delves into the technical details of the study setup, and presents the following topics in order: the simulator experimental setup; the test structure; the scenario rendering information; the design of experiment's details; the participant selection; and finally the data collection process.

SIMULATOR SETUP

In order to realistically simulate semi-AVs and a highway environment, the RiSA²S lab partnered with FKA Prospect Silicon Valley (SV), a subsidiary of the German company FKA (Forschungsgesellschaft Kraftfahrwesen mbH Aachen). The study employed a static driving simulator consisting of a BMW 6 series, a projection wall providing 220-degree horizontal front view, and a split rear-projection wall providing the projection for side and rear-view mirrors.

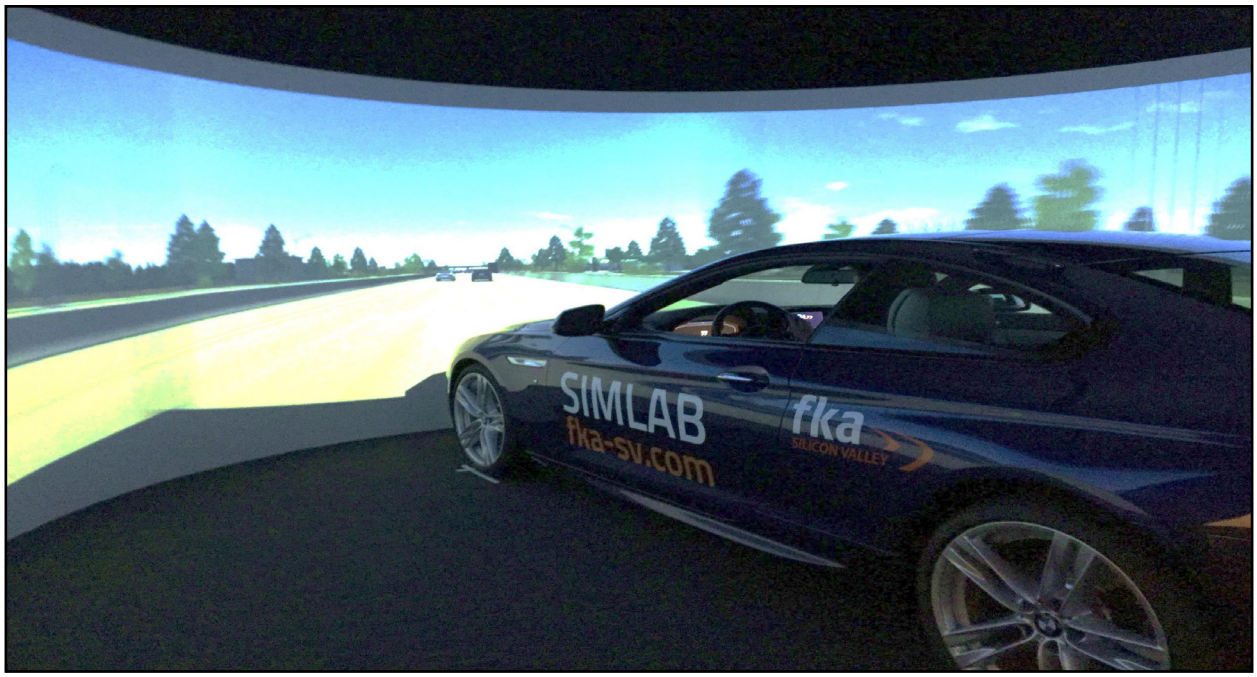

Figure 3. The HITL Simulator Owned by fka Silicon Valley (SV), a Subsidiary of the German Company fka GmbH

Figure 3 showcases a view of the HITL simulator, which is NHTSA-compliant for human-machine interface (HMI) evaluations and is capable of handling both manual control by the driver as well as automated driving. The simulation environment uses the Linux-based simulation framework 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.

While in automated driving mode, the vehicle in the simulator is capable of steering, accelerating and decelerating automatically, and of monitoring the outside environment in order to avoid obstacles and other traffic. In order to simulate driving conditions in compliance with California regulations, we asked and required the drivers to carefully monitor the outside environment, and informed them that a TOR might take place and that, if it did, a warning would prompt them to regain control of the vehicle. Thus, the simulated vehicle was at the border between a SAE Level 2 and a SAE Level 3.

The inside of the vehicle was equipped with a central console with a 10.2 in screen, as well as an analogic dashboard on which the driver could read the speed of the vehicle; both are shown in Figure 4. Drivers could adjust the seats' positions, seat belt's height, side mirrors headings, and rear-view mirrors according to their preference. Figure 4 shows the specific instant at which the visual warning was being displayed on the central console; more details on the specific HMIs will follow later in this section.

Figure 4. Inside View of the Vehicle with Dashboard and Central Console, Shown at the Time of Warning Display

TEST STRUCTURE

Figure 5 schematically shows how each test was structured. The entire experience, from participant greeting to participant dismissal, took place over a duration of 50 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. The control room had an observation window on the simulator environment, as can be seen in Figure 6. The researcher in the control room was in charge of starting the correct test, and monitoring the data logging process. The core of the study (indicated in Figure 5 as “Disengagement Simulation”) lasted up to a maximum time of 30 minutes. Before and after the actual test, however, several other important steps took place; these will be described next.

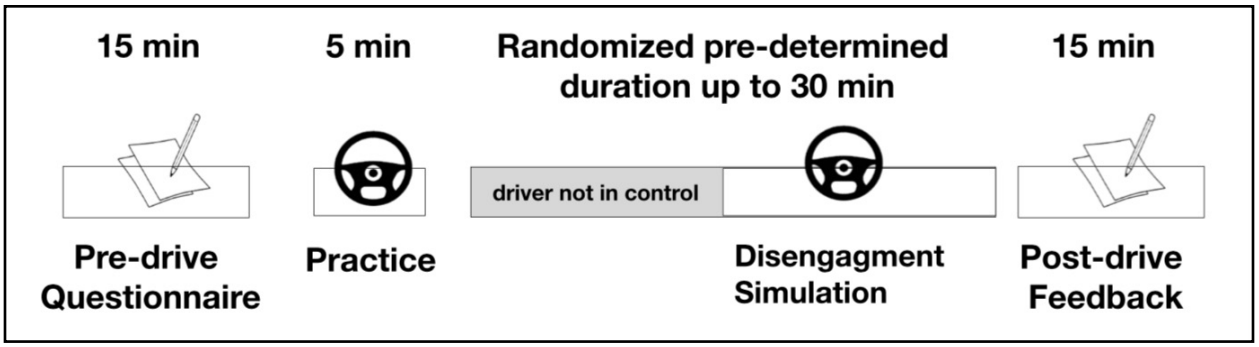

Figure 5. Schematic Representation of the Tests' Timeline for Each Participant, with Expected Times of Duration (Actual Times Varied)

Figure 6. View of the Control Room and Observation Window

Pre-Drive Questionnaire

Before starting the simulation, we asked participants 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 participant's 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 with potential 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).

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 the same track used for the actual study (although the participants did not know that). Although there was a targeted time of 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 in the same order. Specifically, they were asked to change lane, to accelerate and overtake another vehicle, to decelerate and change lanes, and to keep an average speed of 60 mph and follow the road. Once those were executed correctly, the participants were asked whether they were comfortable or not. The baseline S-curve was then executed at the end of the practice phase. Note that the participants were not aware that they were driving the same road they would be tested on. For the entire practice phase, participants drove manually, without assistance from AT.

Disengagement Simulation

Following the practice phase, the actual test began with the car driving autonomously. Before beginning, we gave clear instructions to each participant. Specifically, we told the participants that a disengagement may or may not happen, and that if it did, the vehicle would alert them of the need to regain control; and we instructed them to remain within the same lane of travel. Participants were aware that the test duration was randomized up to a maximum time of 30 minutes. At the time of disengagement, an aural warning repeating the phrase “Danger! Take back control” (human male voice) was provided to the participants until they managed to regain control of the vehicle. Previous HMI literature has suggested that the word “danger” would create a greater sense of urgency than the word “warning”, and that male voice had to be preferred (Bazilinkskyy et al., 2017). The visual warning displayed an exclamation point within a chartreuse yellow triangle, as well as a symbol of hands on steering wheel, shown in Figure 7. Note that participants were not aware of the exact type of warning they would receive.

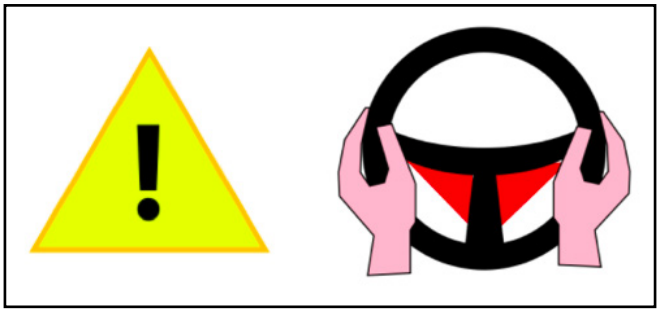

Figure 7. Icon Displayed in the Central Console at the Time of Disengagement

Participants were told only not to touch the radio buttons in the car as they govern the simulation, and to assume a comfortable position, making sure they could at all times monitor the outside environment. Some participants decided to rest their hands on the steering wheel occasionally. An observation form was kept by the researcher in the car to note whether uncommon behaviors were exhibited and whether the person was holding the steering wheel at the time of disengagement.[2]

The disengagement occurred at the end of a straight road, right before the beginning of an S-curve. After the completion of the entire S-curve and the disengagement recovery, the participants were asked to slow down and park the vehicle on the shoulder of the road. This would conclude the simulation test.

Post-Drive Feedback

After participants completed the test, we asked them to fill out a post-drive questionnaire. The post-test questionnaire included four different sections in order to gauge the physical/mental condition of the participants. The first portion investigated situational awareness, asking participants to recollect specific details on locations, speed at the time of disengagement, and time spent within the simulation. A second section on the disengagement experience assessed participants' perception of the success of the recovery, as well as their impression of where they kept their focus/gaze during the simulation. A third section investigated the participants' HMI preferences, whether they considered the aural and visual warnings to have helped or hindered the experience, and their 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.

SCENARIO INFORMATION

Participants executed the test in a closed-track highway-like simulated environment. The track, depicted in Figure 8, consists of four identical-in-shape sections connected by four S-shaped curves, also called 'reverse curves'. We chose the S-curve shape specifically because it allows us to measure and test the quality of the control take-over and the overall performance of the test drivers as explained next. We chose the S-curve for the disengagement locations specifically because it allows for the measurement of the quality of the human driver's recovery: as the vehicle disengages, it continues heading in a straight line, rather than following the road; we assessed the quality of a driver's recovery by measuring the vehicle's drift from the centerline of the road, and the angular difference between the vehicle's heading and the direction of the road. In addition, similarly to the (Naujoks et al., 2017), the driver's performance during their recovery in the S-curve was compared to their initial test drive when they manually drove the vehicle through the exact same S-curve. Their performance during the initial test drive acted as a baseline for which their recovery can be compared to. The disengagement trigger points, marked in Figure 8, were always placed just prior to an S-curve in the track. There was a total of four disengagement points in "invisible" location along that track; only one trigger point was active for each test, depending on the combination of speed and duration tested.

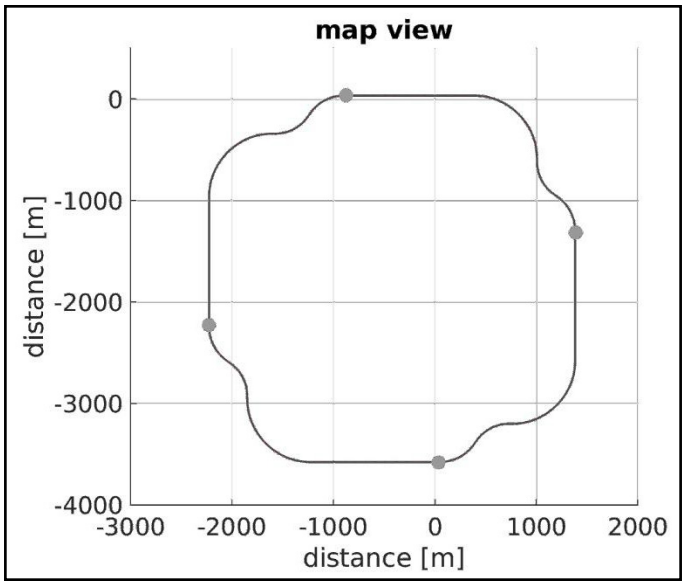

Figure 8. Top View of the Test Track with Highlighted Trigger Points

Note: The flow of traffic is counterclockwise.

The loop obtained by the combination of the four tiles created a closed track of 7.6 miles (12.23 km). The S-curves had radii of 400 meters (1312.34 feet) in order to keep them as realistic as possible while also having a sharp enough curve (based on specs from the US Department of Transportation (DOT) Road Design Manual (Garcia, 2014)). Figure 9 shows the view that the drivers saw, just prior to an S-curve in the track, while Figure 10 shows the top view of an S-curve in the track.

Figure 9. Driver's View of the Simulation Environment

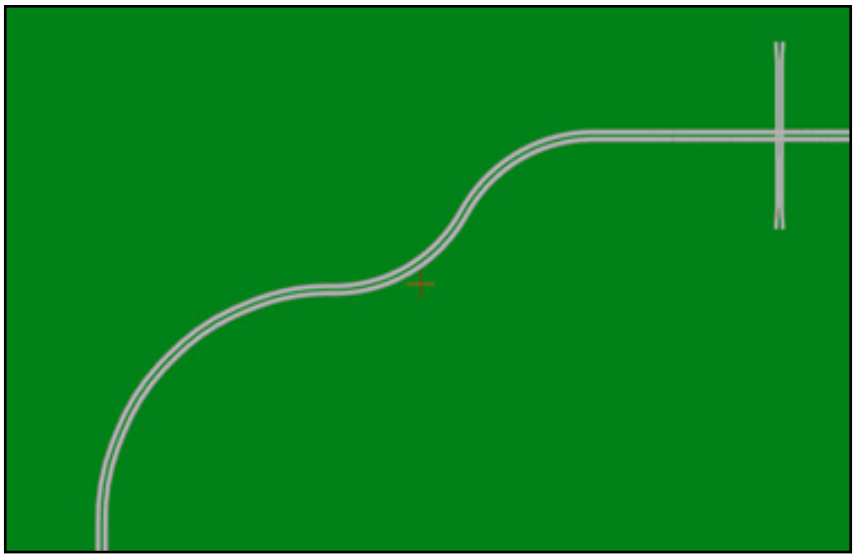

Figure 10. Top View of One of the S-Curves

Even though the four road tiles with the S-curves have the exact same road geometry, each tile had distinct buildings, as seen in Figures 11–14. We made this choice in order to reduce the impression of driving in a loop. Note that the maximum length of the test of 30 minutes corresponded to a total of 4 executions of the loop. Moreover, drivers' situational awareness was also investigated by asking them if they recollected specific buildings or were able to identify the geometry of the road structure; only one participant correctly assessed that the track was a loop.

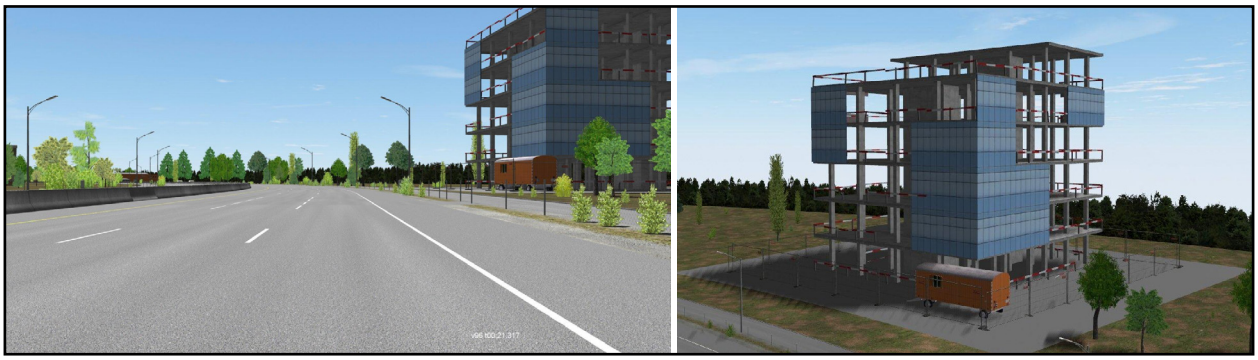

Figure 11. Screenshots of the 'Construction Site' Located Next to the First Disengagement Location

Figure 12. Screenshots of the 'Gas Station' Located Next to the Second Disengagement Location

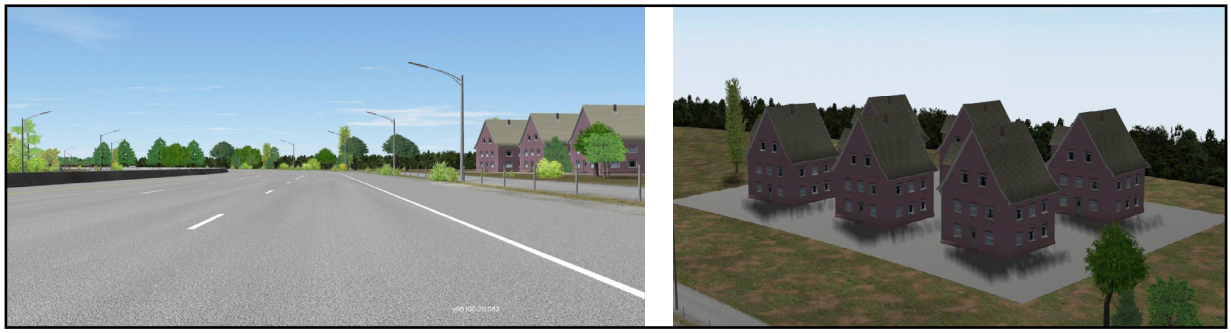

Figure 13. Screenshots of the 'Houses' Located Next to the Third Disengagement Location

Figure 14. Screenshots of the 'Office Building' Located Next to the Fourth Disengagement Location

We did not choose to investigate the effects of traffic density in this study. Traffic density was kept constant for all tests, for a total of 50 vehicles distributed within a 400-meters diameter from the test vehicle. 40% of those vehicles were generated in front of the test-vehicle and another 30% were generated behind the test-vehicle. 15% of the vehicles were generated each on the left and right hand sides of the test-vehicle. Out of the total number of vehicles generated, 60% traveled in the same direction as the test-vehicle, with the remaining 40% traveling in the opposite direction on the other side of the highway divider. Once the vehicles left the 400-meter radius (marked by the outer edge of the yellow circle in Figure 15) they were deleted and were regenerated in an area of 250 to 400 meters (820.21 to 1312.34 feet) around the test vehicle (yellow area in Figure 15).[3]

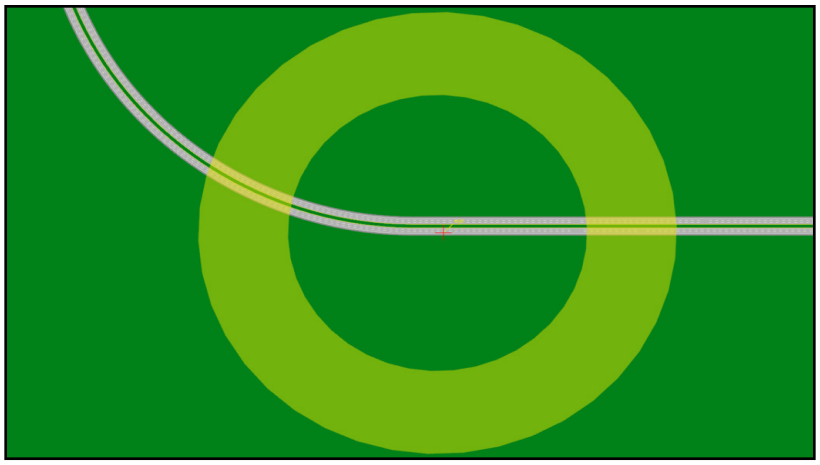

Figure 15. Traffic Generation Area

Note: The red cross represents the test vehicle.

DESIGN OF EXPERIMENT DETAILS

Table 2 summarizes the independent and dependent variables used in the study, and the categories for each. The three independent variables identified for the study are:

1. Age of driver: divided into three levels, from 18 to 35, from 35 to 55, and 55+.

2. Speed at time of disengagement: divided into two settings, high speed and low speed. The vehicle travelled at an average speed of 60 mph during the autonomous driving portion of the test. The actual disengagement happened at either a low-speed setting of 55 mph, or a high-speed setting of 65 mph.

3. Time of the disengagement: the time setting corresponded to the duration of engagement of the autonomous technology. Disengagement times were categorized into three bins, separated by 10 minute intervals, with maximum possible duration being 30 minutes. The AT disengagement occurred during a randomized predetermined one of three time categories, with one disengagement per test. The exact time of disengagement was also influenced by the speed setting used in that test, since disengagements were always triggered at the beginning of the S-curve.

Table 2. Dependent and Independent Variables of the Study

Variable	Type	Number of Categories	Description
Age of Driver	Independent	3	18–35, 35–55, 55+
Speed Setting	Independent	2	Low (disengagement occurring at 55 mph); High (disengagement occurring at 65 mph)
Time Setting	Independent	3	< 10 min, 10–20 min, 20–30 min
Lane Drift	Dependent	N/A	With respect to center line of the lane; maximum lane offset considered
Performance from Baseline	Dependent	N/A	Integral offset comparison between the manual training and the automated test
Response Times	Dependent	N/A	Measured in seconds (continuous). Considered response times are: time to consistent steering input, time to throttle (i.e., acceleration) input, time to braking input

The three independent variables for this study (age, speed, and duration of engagement) were chosen to answer specific operational questions in relation to the use of semi-AVs on US public roads, and that were of particular interest for the specific funding program that sponsored this research: Will there be constraints on the maximum speed the systems can be operated at? Will there be constraints on the maximum time the system can be operated for? Should semi-AVs be sold only to people below a certain age threshold?

The three discretized independent variables allowed us to choose a 3x2x3 factorial design. We executed two full-factorial repetitions, leading to 36 test scenarios with 18 male and 18 female participants (10% of the entire population suffered from nausea, leaving 36 usable data points from the original population of 40; this is described in the participants' selection section).

We investigated the effects of these factors on three dependent variables. The first independent variable listed in Table 2 is raw lane drift, measured as the maximum distance from the center of the vehicle to the centerline of the lane. This quantity is important to consider, as it can be related to unintentional lane departures during the recovery maneuver; considered alone, however, it can be biased, given that a driver's performance in a simulator environment can differ from the actual performance the same driver would have on real roads. The second independent variable listed in Table 2 therefore compares the lane drift obtained during the test to lane drift obtained during the baseline manual practice driving; this is done, not by comparing the maximum distances from the centerline, but rather through a quantity called the "integral offset ratio", which we introduce in the results section. The last dependent variable is the response time to the TOR. Three different inputs are considered here and discussed in the results section: time to consistent steering input, time to throttle (i.e., acceleration) input, and time to braking input.

PARTICIPANT SELECTION

After receiving approval from the Institutional Review Board (IRB) of San Jose State University (SJSU) in relation to human testing, participants were recruited via a flyer posted around the urban SJSU campus. 36 participants completed the study (18 male and 18 female); four additional participants did not complete the study due to motion

sickness. All 40 participants were entered into a random draw to win $100 as an incentive for participating in the study. In order to be selected for this study, all participants needed to have a valid US driver's license and 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. Furthermore, participants were selected by age and gender in order to reflect the age and gender distribution of US licensed drivers as reported by the Federal Highway Administration (FHWA, 2016), shown in Figure 16. This lead to 12 participants between the ages of 18 and 35 (mean group age = 25), 12 between 35 and 55 (mean group age = 46), and 12 older than 55 (mean group age = 60). Gender was evenly split amongst all three age brackets (six male and six female in each age group for a total of 18 males and 18 females).

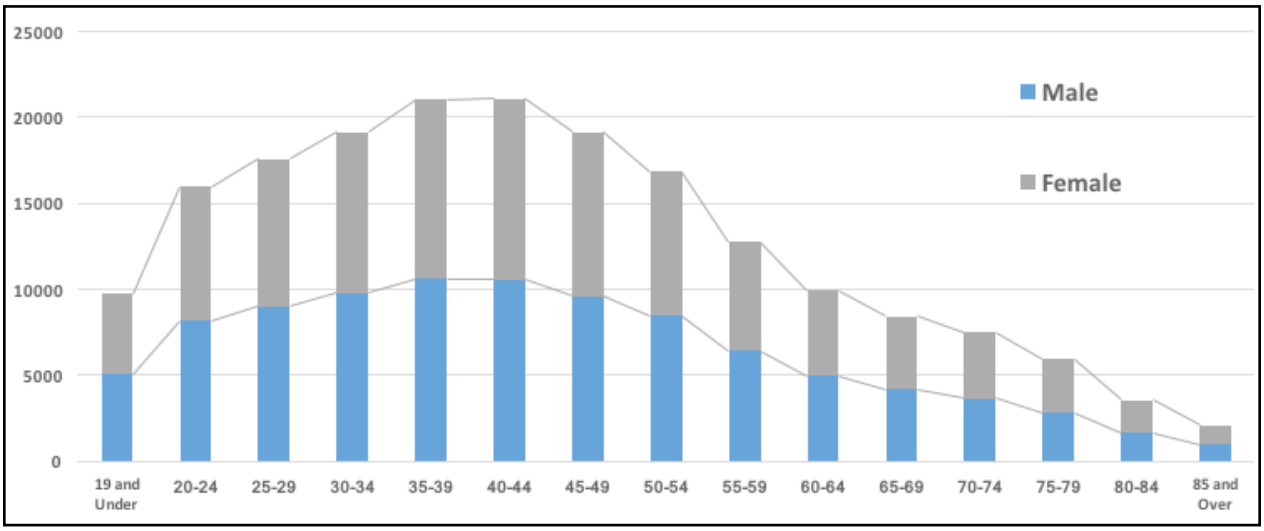

Figure 16. US Age and Gender Distribution of Drivers (in thousands)

Note: Data used for distribution obtained from FHWA (US Department of Transportation, 2016).

Drivers below 18 years of age were excluded from the study due to the necessity of parental agreement at the stage of informed consent collection, and were thus not included in the approved IRB protocol. The category of "19 and Under" from Figure 16 was thus captured by participants in the age of 18 and 19 only. As gathered from the data, such category represents a lower tail in the drivers' distribution. Considering that the age of the oldest participant was 65, this means our tests captured the core of the distribution and excluded both upper and lower tails. The recruitment phase lasted two months and was organized according to an approved IRB protocol.

DATA COLLECTION

The simulator central computer continuously logged the following quantities.

1. Road geometry.
2. Test vehicle heading.

3. Lateral lane offset (offset between the center of the vehicle and the centerline 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 vs. manual).

Road geometry is important for understanding the specific direction of travel, or “heading”, that the road follows. We measured the test-vehicle heading, which is the heading of the vehicle the driver is in, in order to see how divergent it was from the road geometry. Figure 17 illustrates the difference between these two outputs; the angle between these two heading is termed the “angular error” (the dashed grey line represents the center of the lane).

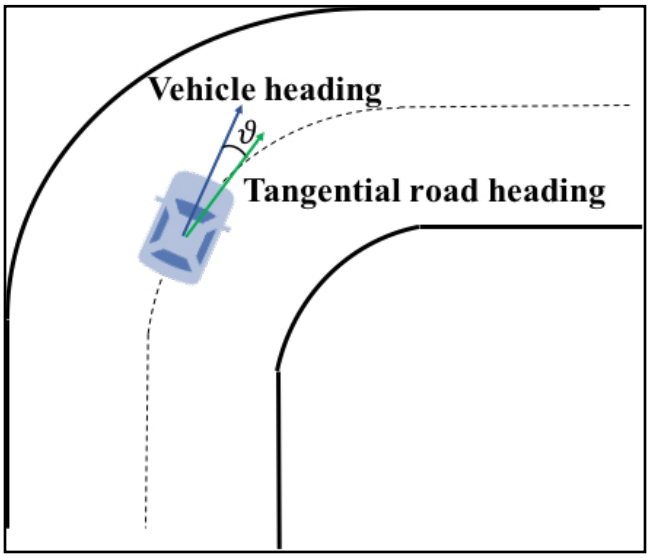

Figure 17. Representation of Angular Error

Lateral lane offset is the distance that the driver let the vehicle drift from the center of their original lane during the disengagement, shown in Figure 18. Lateral offset results from accumulated angular errors that are not corrected by the driver.

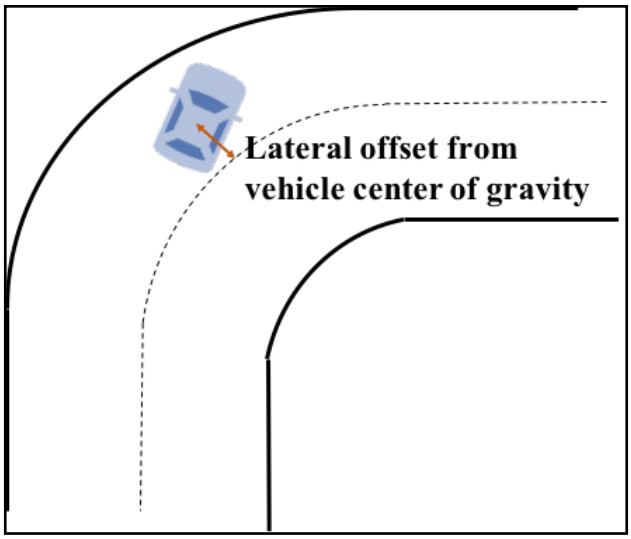

Figure 18. Representation of Lateral Offset

Steering angle is the angle of the steering wheel, given as the angular difference from the neutral position. Brake and throttle outputs are given as percentage, between no depression (0%) and maximum depression (100%).

The output was automatically generated as a csv file. The above listed quantities were measured at all times, both during the practice manual phase and during the actual test.

IV. RESULTS

We divide the discussion of results into three separate sections: (i) the analysis of response times; (ii) the analysis of drift and of quality of control takeover; (iii) human factors results.

RESPONSE TIMES

Table 1 provided a summary of notable literature on the topic of takeover following disengagements in Levels 2 and 3 autonomous vehicles. Before we proceed with the presentation of the results, it is important to clearly define the terminology employed in this work, and address the distinction between the terms “reaction time” and “response time”.

The regulation for AV manufacturers from the CA DMV called for reporting of “the period of time that elapsed from when the autonomous vehicle test driver was alerted of the technology failure and to when the driver assumed manual control of the vehicle” (CA DMV, 2017a). As gathered from Table 1, some authors refer to such data as to “reaction time”, while others employ terms such as “time to takeover” or “response time”. In everyday language, “reaction time” is an all-encompassing term used to refer to how long it takes a person to show a specific behavior (i.e., to react) to a specific stimulus. In the realm of human factors, this usage is partially incorrect. Human factors researchers bring forward the following important distinction between the terms “reaction time” and “response time”. Reaction time is the time between the presentation of a stimulus and the very first measurable activity in the initiation of a response (Wickens, Gordon, Liu, & Lee, 1998): for example, if the doorbell rings, a person’s reaction time may be the amount of time between when the doorbell rang and when they first started to move their eyes in the direction of the door. Response time, on the other hand, is a sum of the reaction time and the time it takes to complete the motor movement for the required response action (Wickens et al., 1998); going back the doorbell example, the person’s response time would be the total time it took them to get up and answer the door. This distinction is important when considering driver reactions to different AV disengagement modes and when assessing human reliability within this domain, with the total response time (or time to takeover) being a more suitable indicator of the actual performance of the human driver in regaining control of the vehicle; what matters for safety is not only whether the driver perceives that a corrective action is needed, but whether he/she also executes it correctly.

Response time within the driving environment typically measures the time it takes the driver to begin their response to an outside stimulus (i.e. the time it takes for a driver to begin depressing the brake pedal after a stoplight turns red). Measuring response time allows for the quantification of driver performance and gives insights into the question of how safely drivers can handle automated vehicle recovery. Moreover, it is important to distinguish which type of response is being recorded or whether other forms of measures are being considered (e.g., reaction time only, and if so how measured). The DMV regulations did not specify whether the input to be considered was steering, throttle, or braking.[4] In order to avoid any ambiguity, in this work we measure three different response times: the time to the first steering input, the time to the first throttle input, and the time to the first braking input.

Selection of Threshold for Collecting the Response Time

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 we 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 (in 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 AT had a lag time of about 10 ms in shutting off all automated outputs.

In order to more accurately select the correct response time, we 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; and (iii) decelerating the vehicle. Visually, this is illustrated in Figure 19.

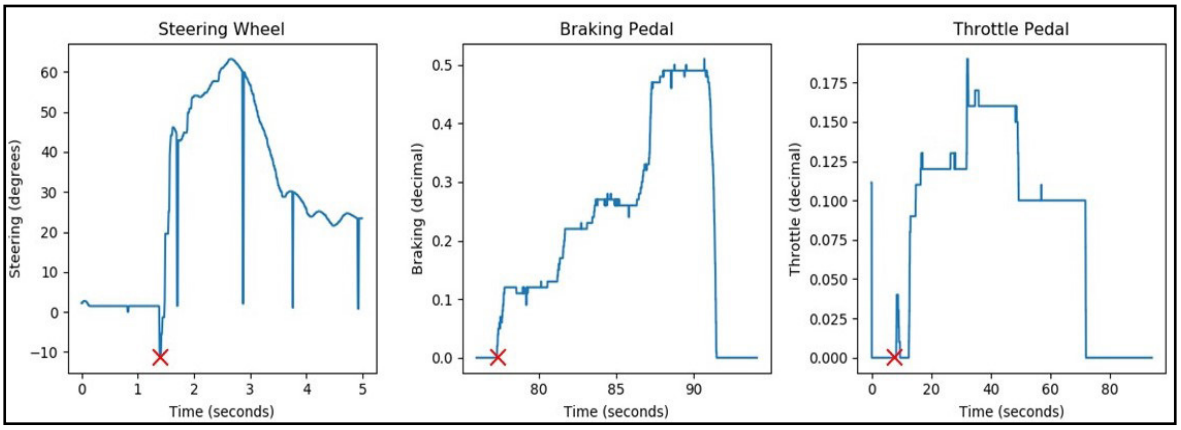

Figure 19. Representation of the Visual Method for Response Time Selection

The method employed is best captured in the left-most picture of Figure 19. The plot represents the steering wheel angle captured by the simulator computer. The S-curve employed for the study started with a turn to the left. Left turns correspond to positive values of the steering angle. As can be seen from the left-most picture, the majority of the recorded response are on the positive side of the abscissa, but two small negative peaks are also recorded before what we 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 we first record a response from the driver to steer the vehicle in the correct direction. Note that in no case did we observe a consistent response in the wrong direction.

We 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).

Response Times Distributions

Figures 20, 21, and 22 provide the probability density functions (PDF) for the three types of response times selected in the study. The non-parametric PDFs were estimated using Epanechnikov kernels and 100 bins. The data analysis was conducted using the R programming language (version 3.5.1) in the RStudio environment (version 1.1.463).

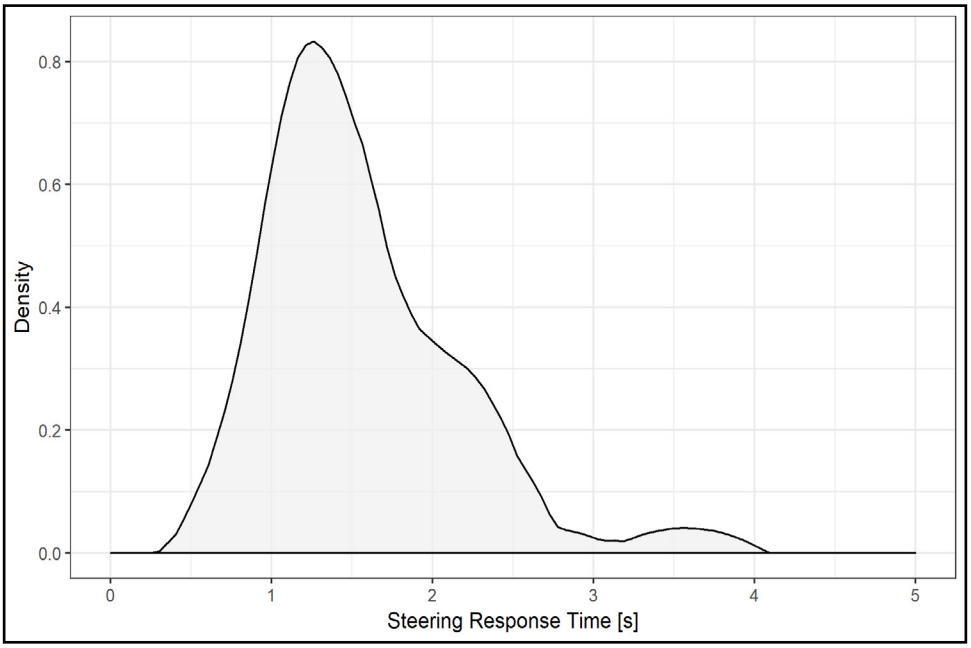

Figure 20. Probability Density Function for Response Time as First Consistent Steering Input

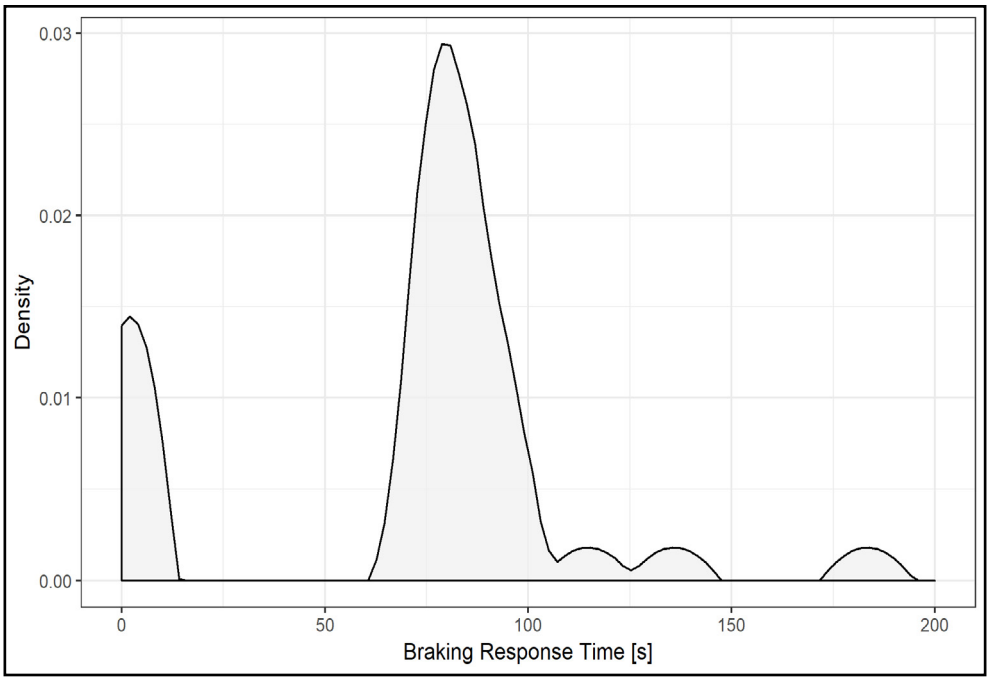

Figure 21. Probability Density Function for Response Time as First Consistent Braking Input

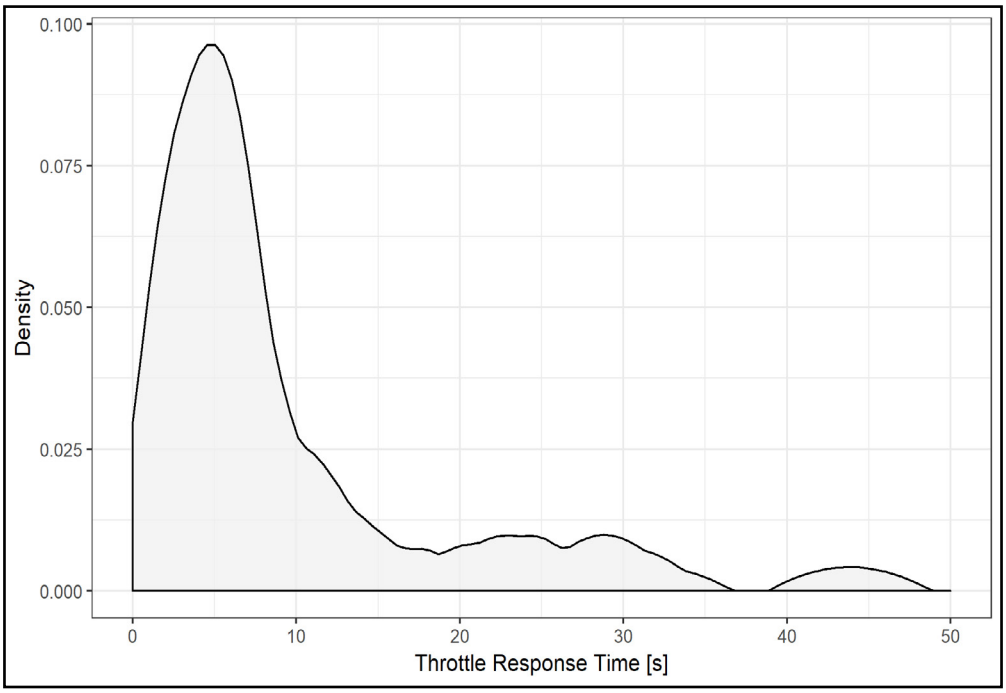

Figure 22. Probability Density Function for Response Time as First Consistent Throttle Input (Acceleration)

The first important thing to note from these figures is the position of the peak for each distribution. Overall, there is one order of magnitude difference between the peak for steering, and the peak for braking. Peaks are located at 1.29 s for steering, two peaks at

5.11 s and 79.06 s for braking, and 4.27 s for throttle. Note that the second peak for braking is spurious, as it refers to the end of the simulation after the completion of the S-curve, when participants were requested to slow down and park the car on the side of the road.

For all participants, steering was the first recorded input. Moreover, note that there was a tendency to accelerate before attempting to brake. 77.7% of the participants resorted to acceleration rather than decreasing the speed of the vehicle right after the disengagement trigger point. This result may seem surprising in light of the geometry used in this study. The disengagement happened at the end of a straight stretch of the road, right before a substantial turn towards the left. We had anticipated that the steep increase in curvature would lead to a braking response. However, we observed the opposite tendency following the disengagement. Peaks of acceleration reached speeds as high as 89.9 mph. A possible explanation could be that, after the disengagement, the vehicle would tend to slow down in the absence of any input due to friction effects. This is because a vehicle that is no longer subject to throttle will slow down due to the air resistance and tire-to-ground friction. Participants could then accelerate in response to such slowdowns. This, however, would not explain the extent of the acceleration observed, and also contradicts the recollection results of about one third of the participants, who wrongly recollected braking as their first response to the disengagement.

Eight participants braked before re-increasing the speed of the vehicle. This cautionary attitude led to the first small peak of Figure 21. Braking pedal usage peaks ranged from 22% to 60% of the maximum vehicle deceleration. Note that, after the S-curve, the vehicle had to come to a full-stop, as participants were asked to park on the right shoulder of the highway.

Response Times: Variable Dependence

In order to investigate the dependence of the response times on the three independent variables used for the study (i.e., age, speed, and time of disengagement), a three-way analysis of variance (ANOVA) was conducted at a 95% confidence level (significance level of 0.05). Tukey's HSD test was used to compare condition means and find the level groupings for each factor. This test measures the "honestly significant difference" between two means. In other words, it is used find means that are statistically different from each other.

The observed factors that significantly impacted steering response time were speed and age (for speed setting $F(1,4)=6.378$, $p=0.0212$; for age $F(2,4)=4.498$, $p=0.0260$). The time of engagement was found to have a marginal significance, with a borderline p-value of 0.056, which did not allow for the rejection of the null hypothesis. While further research with a larger sample will be needed to further investigate the effect of time of engagement, Figure 23 displays the dependence of the steering response time on both the speed setting (high vs. low) and the age group (18–35, 35–55, and 55+).

Figure 23 shows the boxplots that represent the distribution for the steering response time for all combinations of speed and age. Age buckets are located on the x-axis. For each age bucket, the high-speed setting is represented on the left of each age bucket with circle markers, while the low-speed setting is displayed on the right of each age bucket with triangle markers.

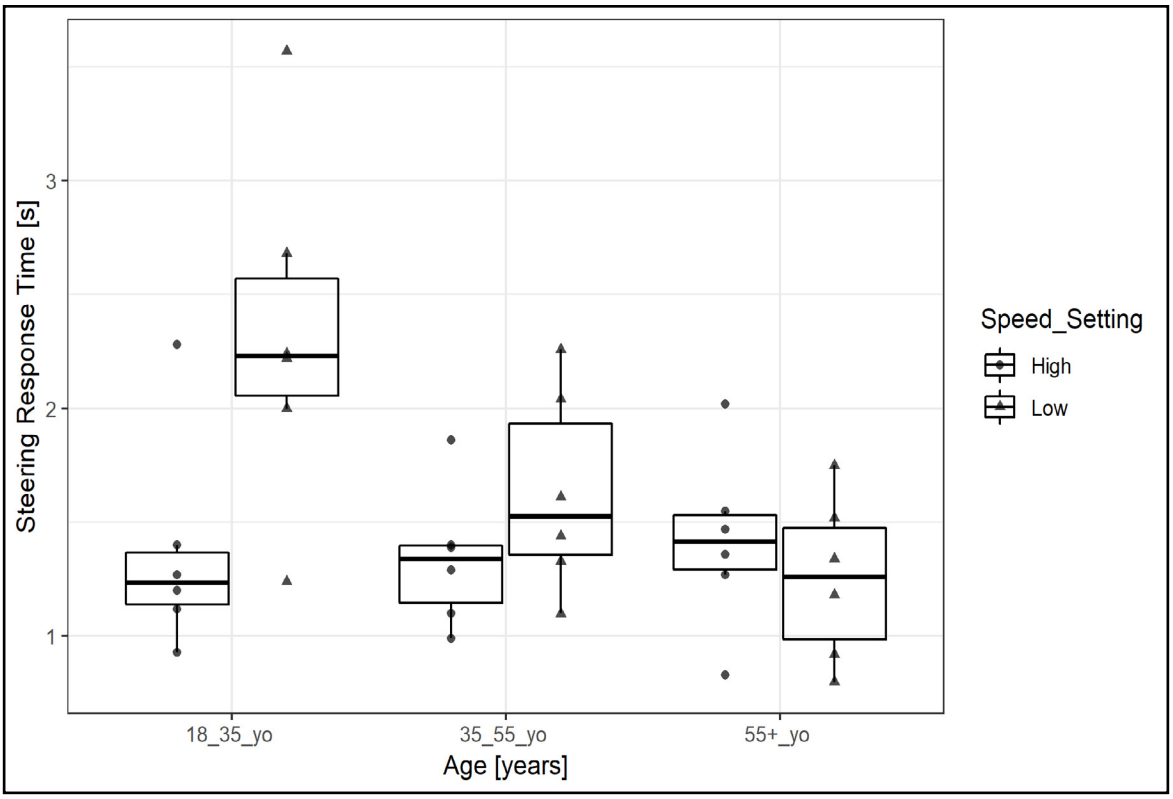

Figure 23. Interaction Boxplots for Steering Response Time as a Function of the Speed Setting and the Age Group

The following trends can be observed:

- At low speeds the response time is decreasing with age. From a high of 2.21 seconds for the younger age group, the response time goes down to 1.35 seconds for the older age group.

- This trend is reversed, although on a smaller scale, at high speeds, with older participants showcasing a slightly higher response time than the younger ones.

The ANOVA showed statistically significant interaction between the speed and age factors ($p = 0.0166$). Note that the performance of the older age group is the one that varies less between the two speed settings, while the spread is a lot more pronounced for younger drivers. This effect will need to be further investigated in future research, but a possible explanation could potentially be that more experienced older drivers may be less affected by changes in speed. It is important to note the presence of outliers for both the combination of young age and low speeds and the combination of old age and high speeds. Tukey's HSD with alpha value at 0.05 also confirmed a statistically significant difference between the mean steering response times of the 18–35 and 55+ age groups.

Neither braking nor throttle response times showed statistically significant variation with the three investigated variables.

Overall, we found good accordance of the results for steering response time with those reported in the literature summarized in Table 1. Good accordance was also found in respect of the “time of first input”, considering that steering was always the first input in our tests (also an intuitive result). Some level of agreement was found with those studies that reported longer response times for throttle engagement compared to steering. The authors were unable to find a detailed analysis of braking habits after disengagement in the current literature.

DRIFT AND QUALITY OF CONTROL TAKEOVER

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

We focus here on two main metrics for the quantification of the takeover performance. The first metric is the maximum lane offset after the TOR and within the first 150 meters of the S-curve. For all participants, drift and erratic driving behavior (steering, throttling, and braking) peaked within the first portion of the curve, allowing us to keep the focus on the first 150 meters after the disengagement trigger. An example is illustrated in Figure 24 for one of the tests showing measured lateral offset and distance in meters on the bottom and feet on the top.

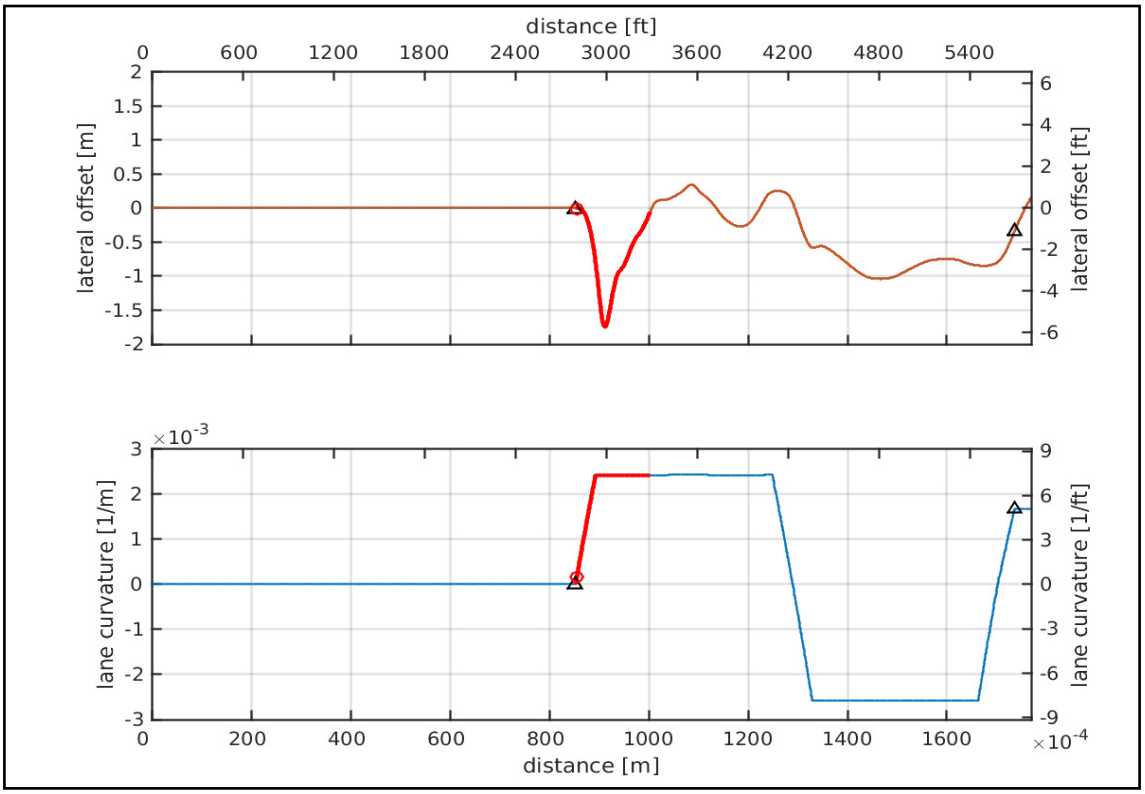

Figure 24. Example Lane Drift and Track Curvature as a Function of Track Distance

Note: Peaks of interest are included within the first 150 meters after the disengagement trigger point for all tests.

The second metric, which we denote by integral offset, encompasses the overall behavior within the first 150 meters (not just the peak) by computing the integral of the car's lateral offset from the lane centerline. 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 the same S-curve 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.

Lane Offset

Figures 25 and 26 summarize the trajectories driven by all participants in their manual drive (Fig. 25) and in their recovery after the TOR (Fig. 26). The "zero" value 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 red 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 right-most lane of a three-lane highway environment. Figures 25 and 26 also depict the respective standard deviations (blue solid lines) and the 95% confidence interval for the observed trajectories (dashed blue lines). Finally, the solid bold black trajectory in each plot describes the mean trajectory, while the thin grey lines are the individual trajectories of each participant.

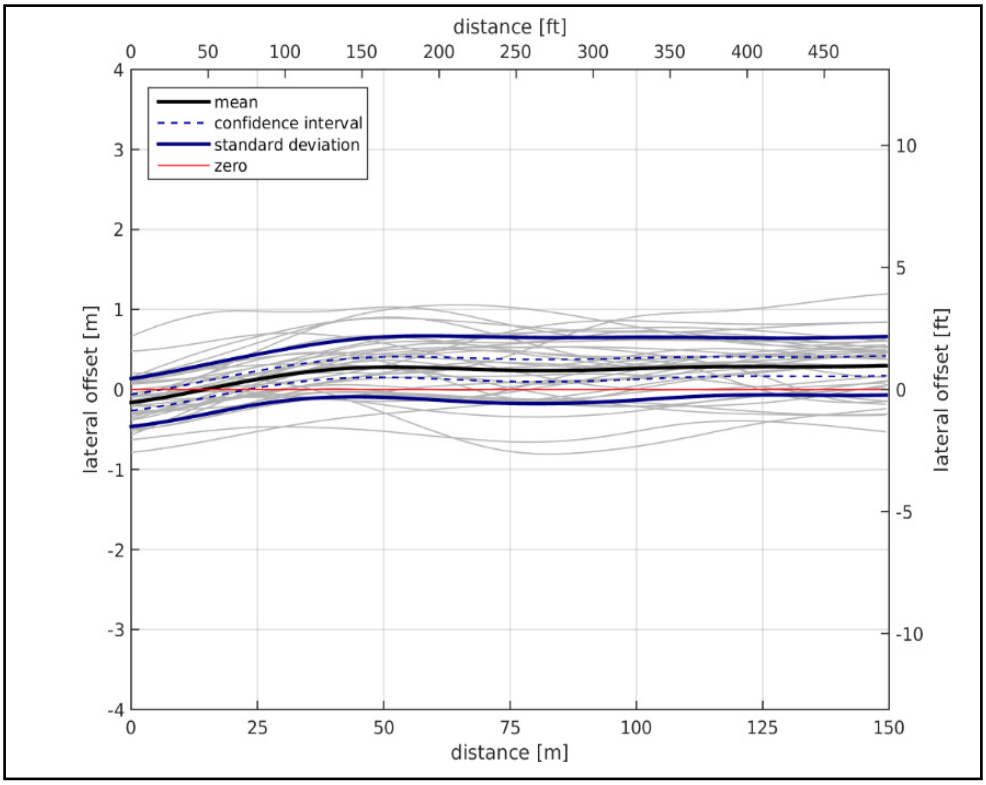

Figure 25. Observed Trajectories During the First 150 Meters (~500 ft) of the S-Curve for the Manual Drive Baseline

Note: Lateral offset from the lane centerline.

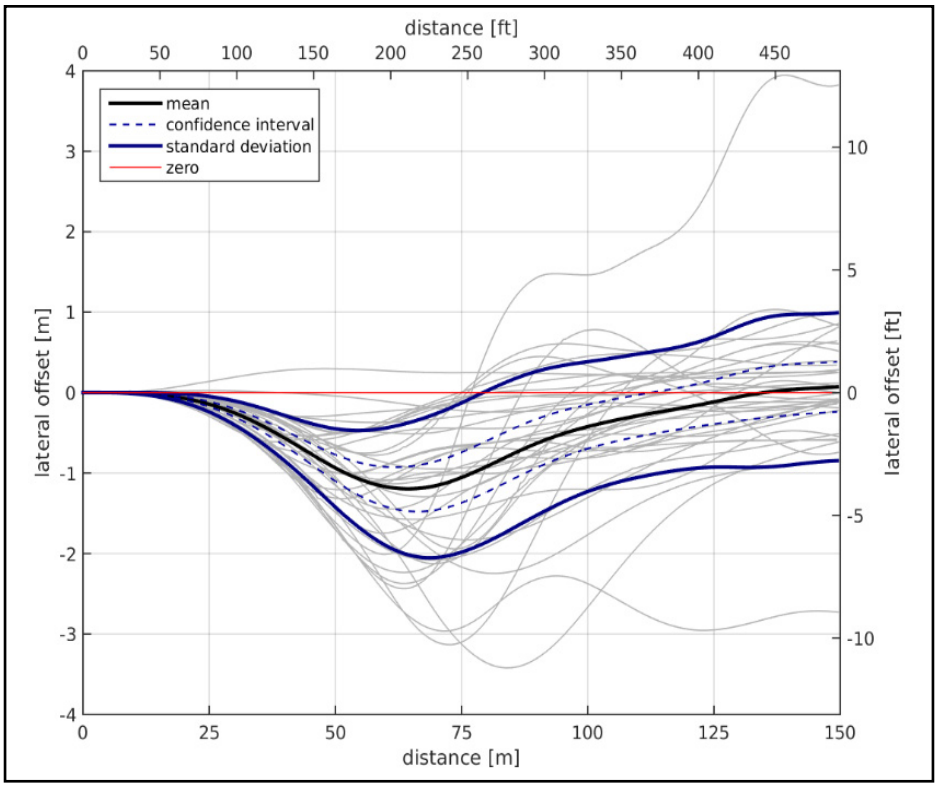

Figure 26. Observed Trajectories During the First 150 Meters (~500 ft) of the S-Curve Following the at Disengagement

Note: Lateral offset from the lane centerline.

At a first glance, it appears evident that results from the manual drive (left) are less scattered and more precise, as expected. Also note that all curves in Figure 26 for the automated test start at zero, as up to the point the AT executed a perfect trajectory with no drift. Figure 27 provides an overview of the distributions of the absolute values of the maximum lateral lane offset, for the manual and automated modes for all 36 subjects. Note that the lateral offset is computed with respect to the center of the vehicle. To determine a possible lane departure of the vehicle, half of the vehicle's width has to be added to the lateral offset reported in all the data shown in this paper. The lane width used for this study was 3.6 meters, with a total vehicle width of 1.9 meters. Considering that the centerline is located in the middle of the lane, this implies that lateral offsets greater than 0.85 meters indicate that the vehicle is crossing a lane marking with the outer (in this case, right) wheels.

Figure 27 highlights a much lower maximum lane offset for manual driving than for the takeover after disengagement (median for manual driving at 0.54 m (within the lane) vs. 1.17 m (crossing lane marking borders) for the recovery after AT disengagement automated test). Furthermore, the data obtained during the manual driving session is less scattered than for automated mode (standard deviation of 0.24 m for manual driving vs. 0.94 m for the automated test). This implies that the population sample achieved similar performances while driving the S-curve during their manual training session, but more diverse responses during the control takeover following the disengagement. However, it is worth noting that some subjects achieved similar results in both their recovery from automation failure and in their fully-manual execution of the S-curve.

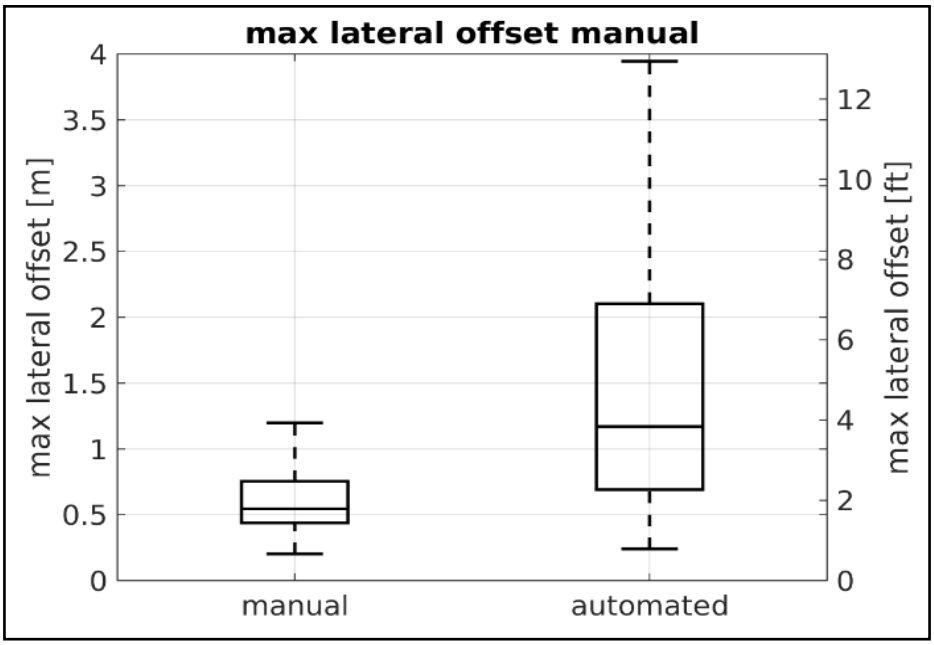

Figure 27. Distribution of Maximum Absolute Lateral Offset for Manual Driving and Automated Tests

Table 3 summarizes the main findings in terms of average values, population minima and maxima, and standard deviations for drivers' maximum lane offsets.

Table 3. Summary of Statistics for Lane Offset, N = 36

Variable	Value for automated test (after the TOR)	Value for manual driving
Average Maximum Lane Offset	1.45 m	0.60 m
Median Maximum Lane Offset	1.17 m	0.54 m
Minimum Observed Maximum Lane Offset	0.24 m	0.20 m
Maximum Observed Maximum Lane Offset	3.94 m	1.20 m
Standard Deviation for Maximum Lane Offset	0.94 m	0.24 m

Non-parametric probability density functions for maximum lane offset for both manual and automated performance were estimated using Epanechnikov kernels and 100 bins. The probability density functions (PDFs) are shown in Figure 28. Table 3 shows the narrow peak for manual driving results, compared to the high dispersion of the distribution for recovery after automation failure. While means of the two distributions are close, chi-squared testing shows independence of the two variables ($p = 0.347$ for $X^2(200)$).

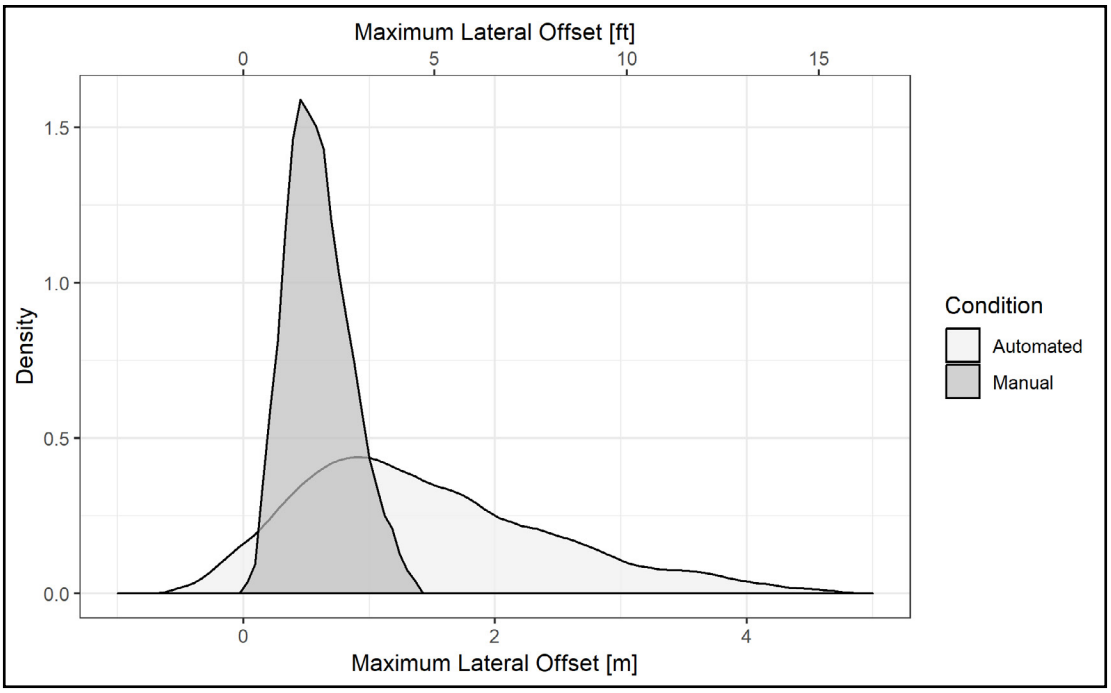

Figure 28. Probability Density Function for Maximum Lateral Offset in Conventional Driving Simulation Versus After Recovery from Disengagement

We conducted a three-way factorial ANOVA in order to understand the interaction between the investigated independent variables and the participants' drift performance. The speed setting [$F(1,4) = 19.293$, $p = 0.000351$] and, to a lesser effect, the time setting [$F(2,4) = 3.668$, $p = 0.0461$] both had statistically significant effects on maximum lane offset at the $p<0.05$ level. The age group did not significantly affect lane offset. Figure 29 provides an overview of the interaction plots related to drift performance as a function of the speed and time of disengagement.

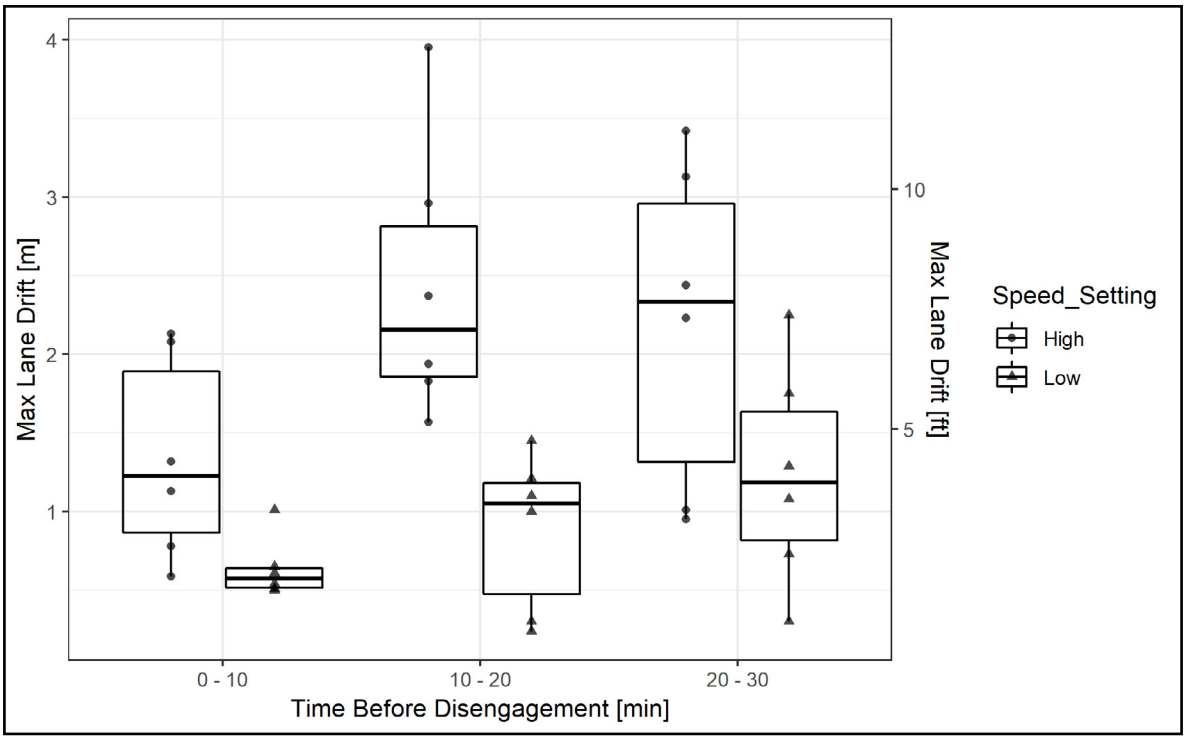

Figure 29. Interaction Boxplots for Maximum Lateral Offset After Disengagement Recovery as a Function of Speed Setting and Time Of Disengagement

Figure 29 highlights the impact of the speed setting on maximum lateral offset. The increase of speed of the AV by 10 mph between the two settings more than doubled the average maximum offset (from 0.91 meters vs. 1.99 meters). This effect is clearly visible for all time bins investigated. Higher durations of engagement led to an increase in drift for the sampled population. This trend can be observed for both speed settings in Figure 29. Unexpectedly, we found that age dependence was found to be not statistically significant ($p >> 0.05$ at 0.3399). A surprising result was that the older group of participants (55+) performed the best at the low-speed setting, and performed comparably to the other age groups at high speeds. Tukey's HSD comparison of means was also executed, with statistically significant results only for the time of disengagement factor, when comparing the lowest setting (0–10 min) to the highest setting (20–30 min).

The minimum observed maximum lane offset was of 0.24 meters, achieved at a low-speed setting, for the second time range of 10–20 minutes, by a female participant over 55 years of age. The maximum observed maximum lane offset was of 3.94 meters, achieved at a high-speed setting, also for the second time range of 10–20 minutes, by a female participant in the 35–55 age category.

Integral Offset – Comparison to Baseline

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 his/her own manual baseline. To compute this metric, we 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 trajectory of Figures 25 and 26 forms with respect to the lane centerline (i.e., the area between the red line and a thin grey line, in Figures 25 and 26). Figure 30 provides an overview of the integral offset distributions for both the manual and automated performance for the 36 tests. The same conclusions highlighted for Figure 27 apply. However, outliers were observed for this metric (shown as "plus" marks in Figure 30). In the case of the automated performance, one of those outliers was due to a lane change to the middle lane within the first 150 meters of the S-curve.

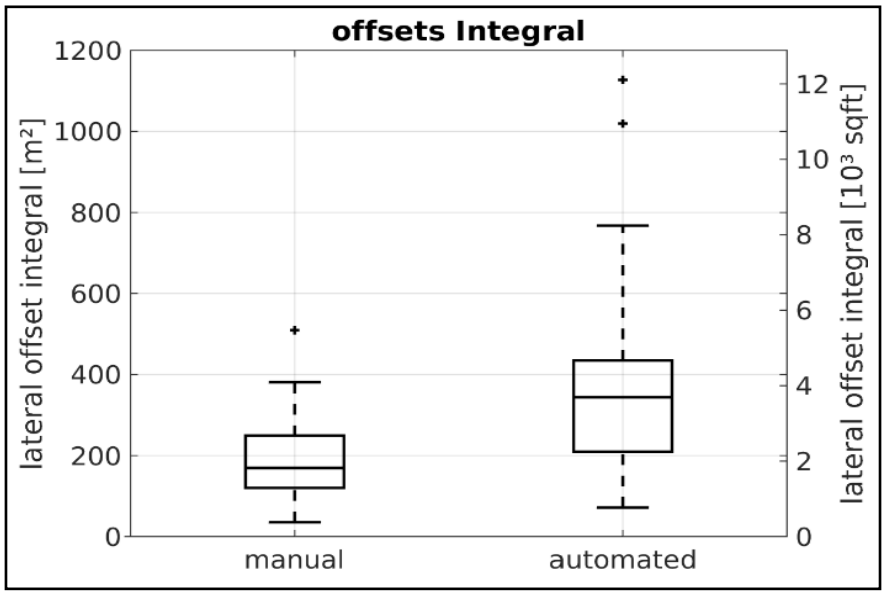

Figure 30. Distribution of Integral Offset for Manual Driving and Automated Tests

Table 4 provides a summary of the main findings in terms of average values, population minima and maxima, and standard deviations for the drivers' integral offsets.

Table 4. Summary of Statistics for Integral Offset, N = 36

Variable	Value for automated test (after the TOR)	Value for manual driving
Average Integral Offset	380.11 m²	195.79 m²
Median Integral Offset	344.08 m²	169.40 m²
Minimum Observed Integral Offset	71.37 m²	35.49 m²
Maximum Observed Integral Offset	1127.73 m²	509.33 m²
Standard Deviation for Integral Offset	239.72 m²	102.23 m²

We computed the integral offset ratio by dividing the integral offset of the automated test by the integral offset of the manual drive. An integral offset ratio of 1 thus implies an overall similar performance along the first 150 meters of the S-curve for drift between conventional driving mode and recovery after disengagement. Values higher than 1 signify that performance is worse during recovery after disengagement, than during conventional driving; these are the values that we expected to see. Values lower than 1 signify that participants performed better after recovery from a system's failure than during conventional

driving modes.[5] The dimensionless ratio is adopted for ease of interpretation instead of a difference, which would have square feet units. Figure 31 shows the non-parametric PDF for the integral offset ratio. The peak is located at 1.278.

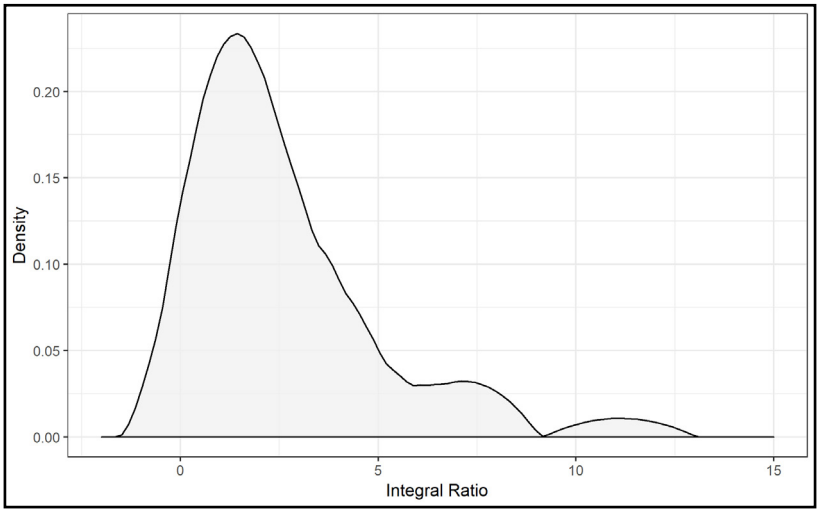

Figure 31. Probability Density Function for the Integral Offset Ratio

We executed a three-way factorial ANOVA for the integral offset ratio. The results of the analysis indicate a statistically significant effect for speed [F(1,4) = 4.484, p = 0.0484), once more, and marginal significance for the age factor [F(2,4) = 3.529, p = 0.0509], with older participants showing lower ratios than younger participants (overall decrease of the ratio, i.e., improved performance, of 64% for the oldest group over the youngest group). Tukey's test confirmed a statistically significant change in the mean between the younger age group and the older age group for the integral offset ratio (p = 0.0177, significance at 95%).

Figure 32 provides an overview of the interaction plots related to the integral ratio as a function of the three investigated independent variables.

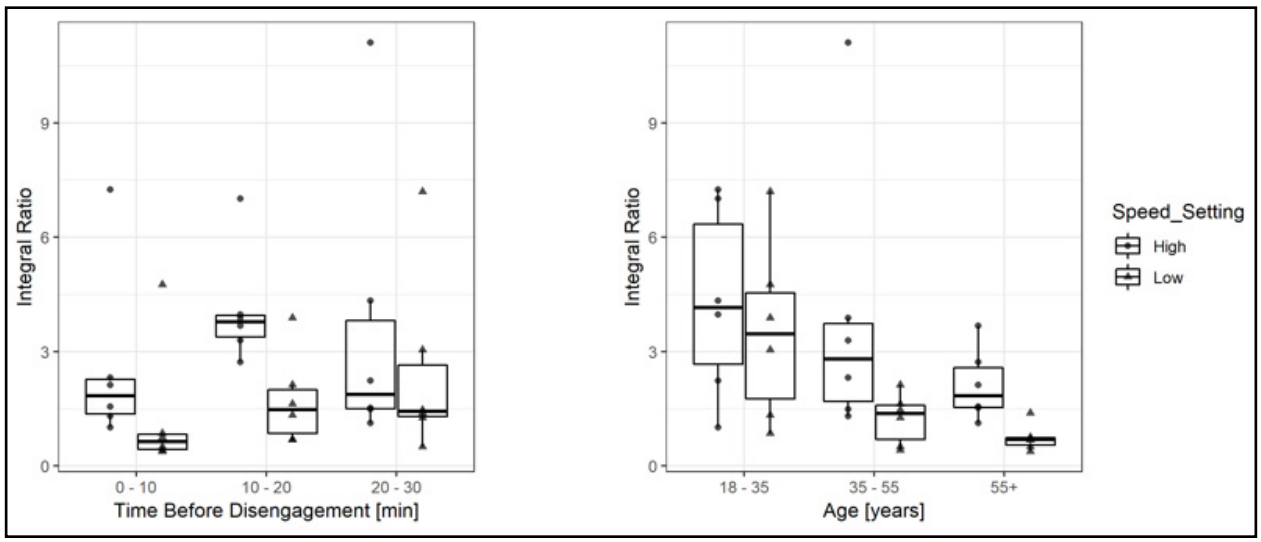

Figure 32. Interaction Boxplots for Integral Offset Ratio as a Function of the Independent Variables

Eight participants out of 36 (22.2%) performed better in their automated test than in the manual conventional drive. In all eight cases, the test was executed in the low-speed setting. Furthermore, for one of the eight cases, the integral offset computed for the manual drive was considered an outlier (this test achieved the maximum integral offset of 509.33 m^2, equal to a 260% increment on the observed average). In terms of age, five were 55+, two were between 30 and 55 years, and one was in the lowest age group. Five of the subjects were in the 0–10-minute range, two in the 10–20-minute range, and one was in the 20–30-minute range. Five subjects were male and three were female.

Similar to what was observed for Figure 29, high-speed settings led to a decrease in performance, with the integral offset ratio going from 1.82 for low speeds to 3.44 for high speeds (an increase of 89%). The disengagement time did not show statistically significant trends, and the authors believe that further refined testing would be needed to detect any clear trends for this variable.

Overall, the minimum observed integral offset ratio was 0.41 (after rejection of the outlier for the automated test), achieved at a low-speed setting (55 mph), for the first time range of 0–10 minutes, by a female participant in the 35–55 age group. The maximum observed integral ratio was of 3.97 (also after rejection of an outlier for the manual test, who matched the centerline of the S-curve within the range of centimeters), achieved at a high-speed setting (65 mph), for the second time range of 10–20 minutes, by a female participant in the 18–35 age category.

HUMAN FACTORS

In addition to executing the core driving testing phase, we asked participants to fill out a pre-test questionnaire as well as to provide post-test feedback. This section summarizes the most notable results obtained from the surveys (which were completed in their entirety by all participants), which we divide for ease of presentation into the following categories: (i) situational awareness; (ii) perception of success; (iii) subjective measures (emotional and physical response); and (iv) HMI preferences.

Situational Awareness

The post-test feedback queried the participants' ability to correctly recollect (if shown a measurement) or to estimate, the speed of the vehicle, the time spent in the simulation environment, their response to the disengagement (in terms of first used input), and their gaze. focus area.

Accuracy Definition

In order to assess the goodness of such measures and thus be able to analyze and interpret situational awareness results, we used a performance measure termed "Accuracy", commonly used in Machine Learning for results with binary outcomes (Zhu et al., 2010). It is essentially a fraction of the predictions/answers that a model gets correct, obtained through the following equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy is calculated by summing the true/correct responses (true negatives (TN) and true positives (TP)) and dividing them by the total number of responses, including false/ incorrect ones (false negatives (FN) and false positives (FP)). An answer is categorized as true or false depending on whether the answer of the participant matches the actual measurement from the simulation. We thus have for instance “True Positives” when a positive answer from the participant matches a positive measured answer; conversely a “False Positive” would be obtained when a positive answer is provided by the participant, but the measurement is actually negative. We used the Accuracy indicator to quantify the quality of the participants’ recollections of the speed at time of disengagement, as well as their perception of success in control takeover after disengagement, as explained next.

Recollection of Speed at Time of Disengagement

After the test completion, we asked participants to report the speed of the vehicle at the time of the simulation disengagement, and whether their recollection was based on an actual reading from the digital speedometer placed in the vehicle dashboard. Their responses were evaluated to be correct if they fell within a threshold of +/- 2mph of the actual speed. After comparing the drivers’ numerical answers against the actual speed in their test, the population’s accuracy was computed. Table 5 that the overall population’s accuracy of speed recollection was 55.5%.

Table 5. Overall Accuracy for Speed Recollection

	TP	TN	FN	FP	Accuracy
Occurrence	18	2	2	14	55.5%

Figure 33 shows the speed recollection accuracies of the subpopulations, categorized according to the values of the independent variables. Age did not show a statistically significant effect on accuracy. Speed of the vehicle did have a significant effect, with higher accuracy at low speeds. Time of disengagement also had a significant effect, with accuracy being lowest for medium durations of engagement.

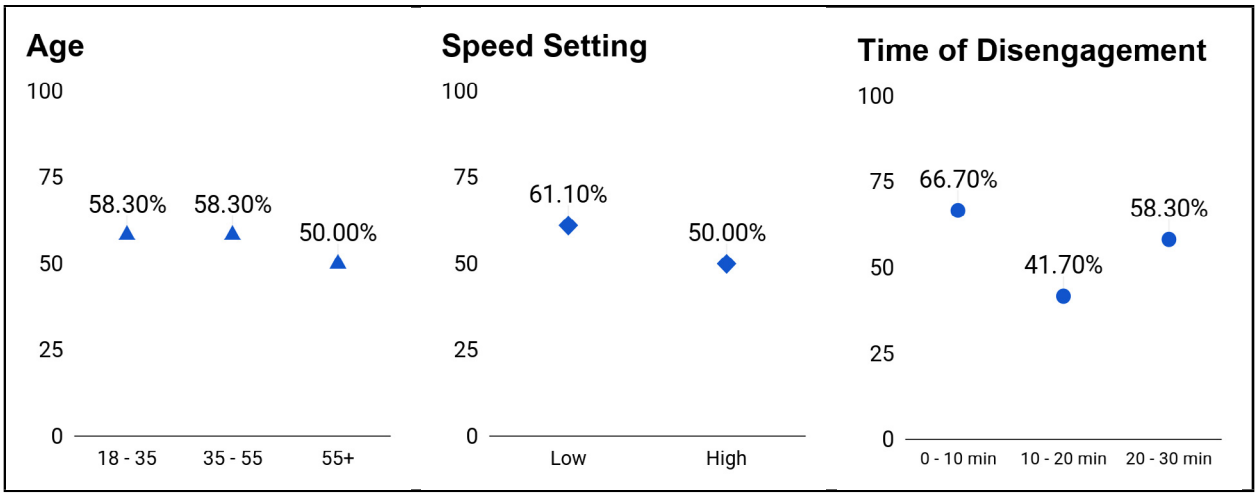

Figure 33. Accuracy in Speed Recollection as a Function of the Investigated Factors

Recollection of Time in the Simulation

We also asked participants to estimate the duration of the simulation. We compared their reported times to the actual simulation times, counting responses as correct if they fell within a threshold of +/- 5 minutes. Additionally, we classified responses as 'Spot On' if their answers had a difference of <1 min from the actual time. The overall test results are summarized in Table 6, indicating that participants were more likely to overestimate the test duration than to underestimate it or be spot on. An examination of factors' effects showed that the middle aged group had the highest "within threshold" estimation percentage, and that shorter test durations led to less precise estimations.

Table 6. Summary of Results for Time in Simulation Estimation

Answers	Occurrence	Percentage	Average Error [min]
Underestimated	7	19.4%	- 4.02
Overestimated	26	72.2%	7.09
Spot on	3	8.3%	0
Within threshold	**17**	**47.2%**	N/A

Figure 34 shows the actual times (blue) and reported times (orange) in the simulation for each participant, ordered by increasing actual time. Figure 35 shows the accuracies of the subpopulations, categorized to the values of the independent variables. Higher speeds and the middle age group showed a better recollection than did the other values of their respective variables. The shorter time duration had the lowest accuracy in time recollection.

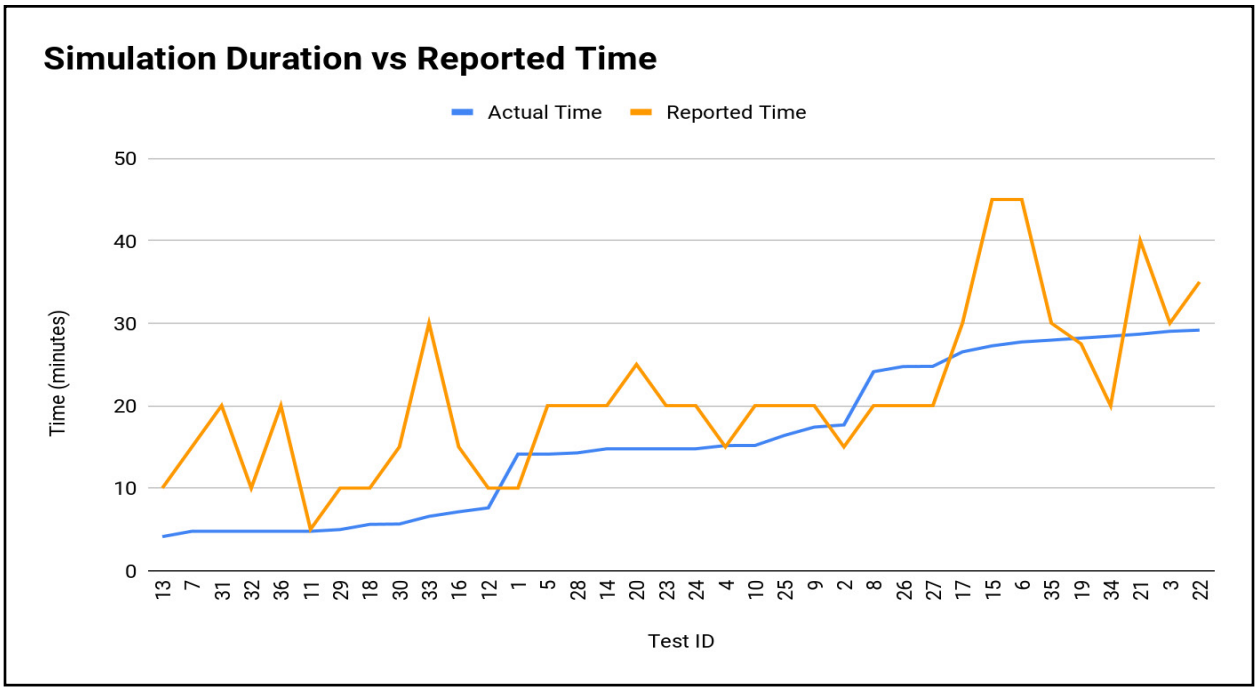

Figure 34. Reported Time in Simulation by Participants Compared to Actual Time in the Simulation

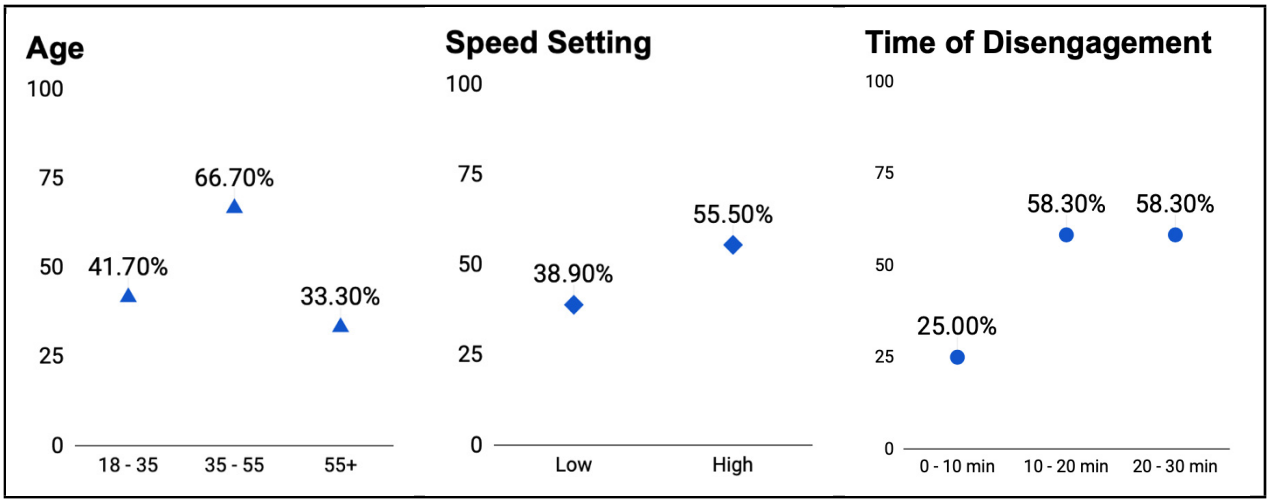

Figure 35. Percentage of Participants Within the Threshold in Time Recollection as a Function of the Investigated Factors

Recollection of First Input at Disengagement

We asked participants what input they first provided as a response to the disengagement. As seen in the previous section, steering was always the first input and we thus analyzed the accuracy in terms of seeing which participants correctly assessed whether throttle or braking was employed first. Only 22% of the participants actually resorted to braking before throttle as a response to the disengagement and to the change in road curvature. Of the 78% that resorted to acceleration first, 32% of them (i.e. 25% of the population) incorrectly

recalled braking to be their response to the disengagement event, when in fact, they had accelerated. This is an important factor to consider, given the prominent effect that higher speeds had on increasing drift and the likelihood of unintentional lane departures.

Gaze Before Disengagement

We asked participants to rank their gaze focus level, from 1 (lowest values) to 5 (highest values), on different parts of the surrounding environment, i.e. outside the vehicle, inside the vehicle and other locations. The overall gaze at different parts of the simulation are represented as weighted averages, shown in Figure 36. Overall, areas outside the vehicle had the highest gaze levels with the front being the highest, (75% of the participants ranked their gaze as 4 or 5), and other locations including inside the vehicles had considerably lower levels. This is an important point to consider when designing human-machine interfaces, which we discuss shortly.

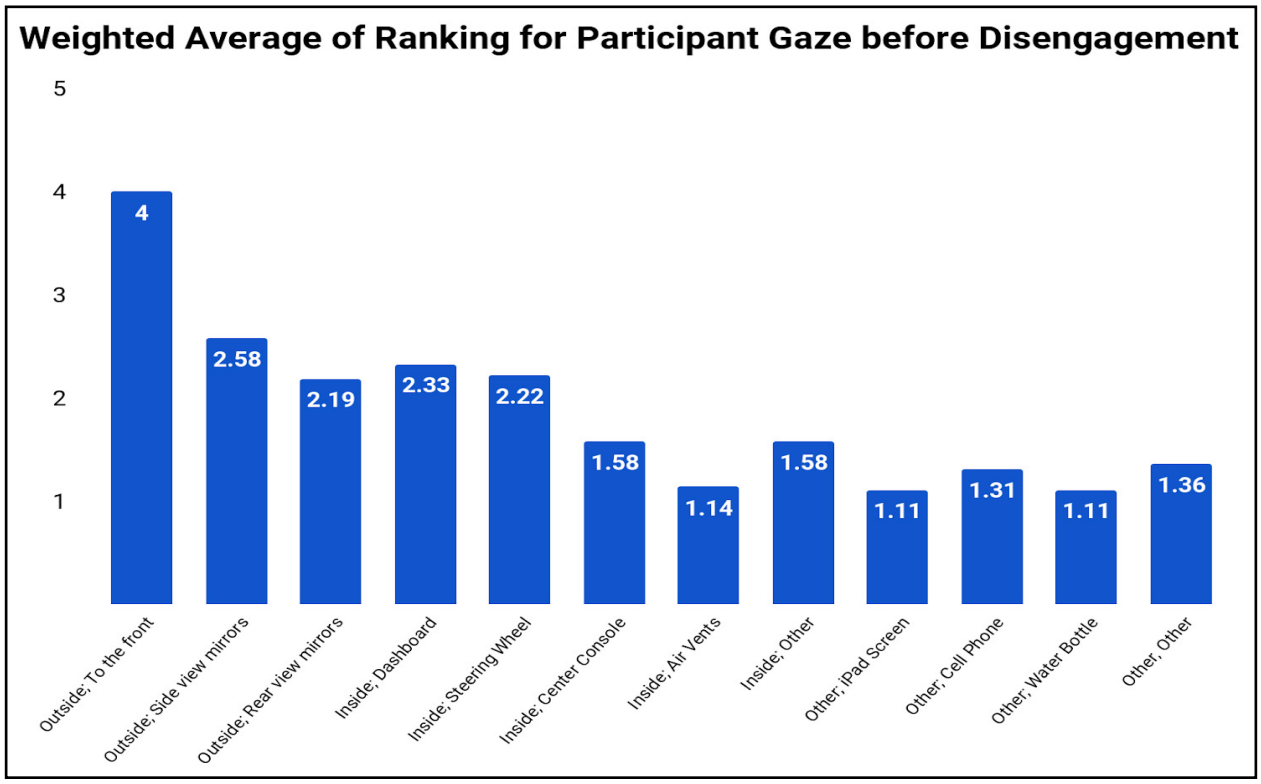

Figure 36. Participant Ranking of Gaze Focus Levels for Different Locations of the Surrounding Environment Before Disengagement Shown as Weighted Averages

Perception of Success

We studied participants' perception of success in control takeover by comparing a binary option of success (yes vs. no) indicated by participants to a binary measure of drift (remained within the lane vs. unintentional lane departure). Table 7 shows the population accuracy, and indicates overconfidence in the quality of the control takeover by the majority of participants, with only one participant recognizing the failed recovery attempt.

Table 7. Overall Accuracy for Recovery Success

	TP	TN	FN	FP	Accuracy
Occurrence	11	1	0	24	**33.3%**

Figure 37 shows the success perception accuracies for the subpopulations, categorized according to the values of the independent variables. The highest accuracies for each independent variable were observed for older participants, for lower speed settings, and for short durations of engagement (which tended to have lower drift) respectively. Older aged participants had a higher accuracy (50% vs 25%) in recalling their recovery after the disengagement. Low speeds and lower test durations also led to better recollection accuracy (50% and 58.3% respectively).

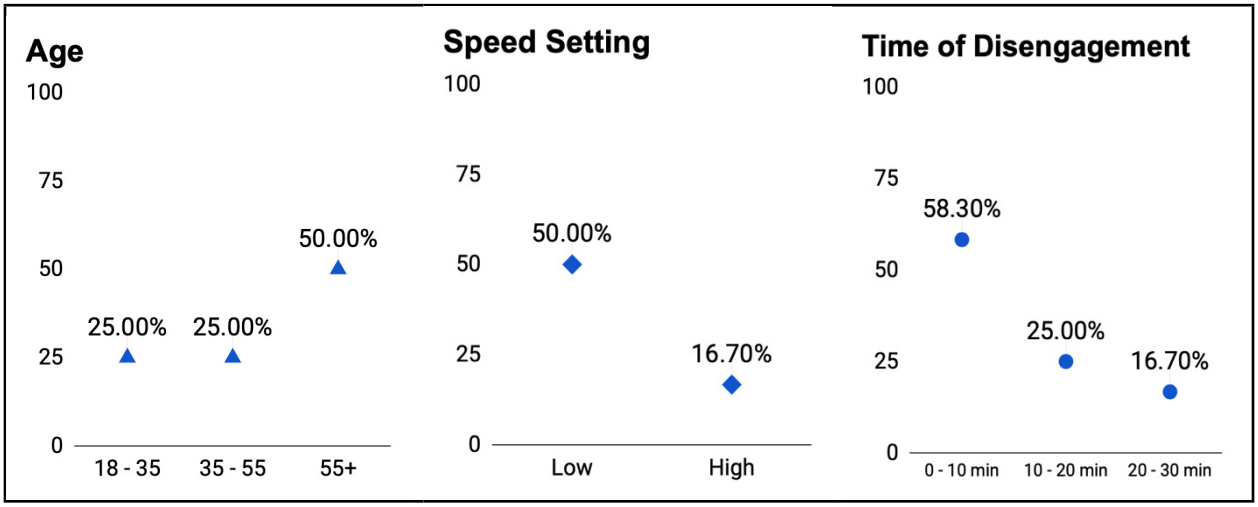

Figure 37. Accuracy in Success Estimation as a Function of the Investigated Factors

Subjective Measures – Emotional and Physical Response

The surveys included queries on the participants' emotional state during the simulation. Figures 38 and 39 summarize the key findings in relation to trust, fear, and nervousness, as well as to changes of trust in the technology after the participants' involvement with the study. Fear and nervousness levels were higher at high speeds than at low speeds. Low speed also led to higher levels of overall trust. Nervousness was higher for the older age group than for the younger one.

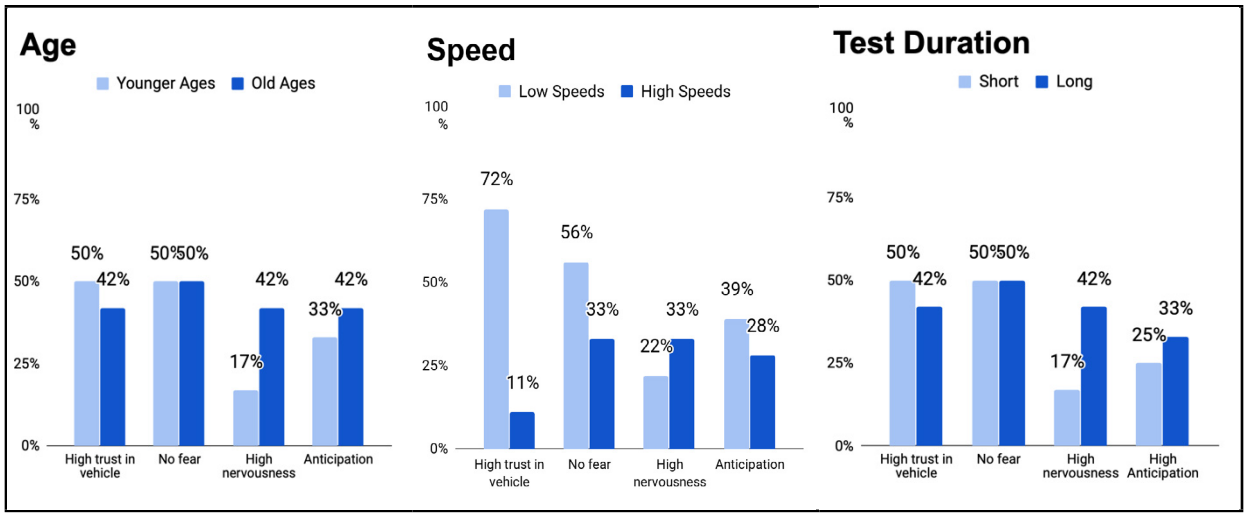

Figure 38. Summary of Main Findings in Relation to Trust, Fear, Nervousness and Anticipation as a Function of the Investigated Factors[6]

Moreover, the older age group showed a greater change in trust level than younger ages. Similarly, high speeds showed greater changes in trust level than low speed, with a higher percentage of decrease in trust at higher speed than low speed. Figure 39 summarizes the participants' change in trust levels, either as an increase (shown in blue) or a decrease (shown in orange) from before to after the test executed for this study.

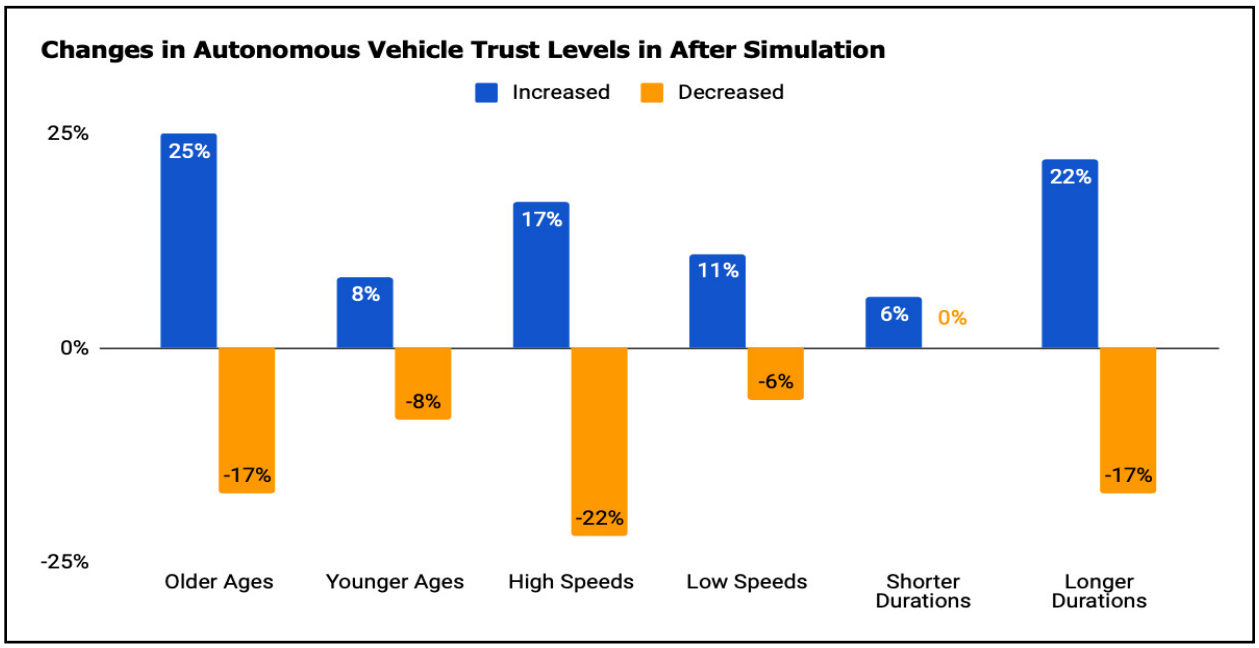

Figure 39. Changes in Level of Trust in the AV Technology (Before vs. After the Experiment) as a Function of the Investigated Speed and Age Groups[7]

With regard to physical effects, we asked participants to rate how nauseous they felt during the simulation, and whether nausea affected their ability to take control of the vehicle following the disengagement. 94% of the participants felt some level of nausea

during the test, ranging from mild to moderate. Figure 40 summarizes the findings and shows that 31% of the participants expressed that nausea affected their driving. 81% of the people who said it somewhat affected their driving did not successfully recover from the disengagement. Moreover, the one participant who expressed that it affected their driving significantly had a maximum lane drift in the higher range (3.13 m), while the average drift of all the participants who expressed that nausea affected their driving was 1.59 m, which was slightly higher than the average maximum drift of all participants (1.45 m). Nausea is known to possibly create bias in the results of simulated studies, and this is one of the reasons why we opted for setting up the novel metric of the integral offset ratio, so that each participant would have their own simulated (and possibly affected by queasiness) baseline to which to compare their performance.

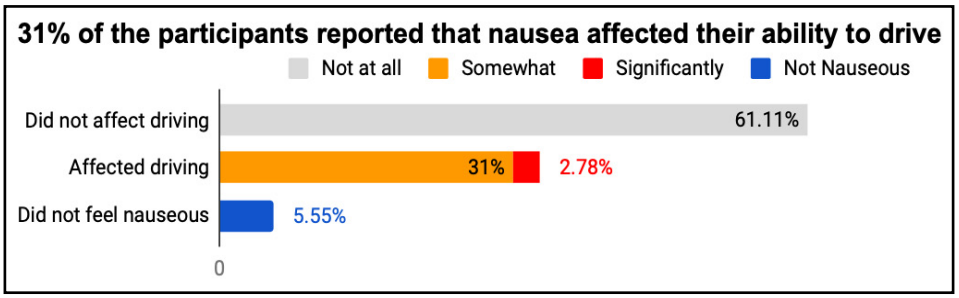

Figure 40. Participant Response on Nausea During the Simulation

HMI Preferences

The HMI interface employed for the study was presented in the Methodology section, and it included a male voiced warning “Danger! Take back control” repeated until control was achieved by steering the vehicle. Moreover, a visual warning in the central 10.2-inch console displayed an exclamation point within a chartreuse yellow triangle as well as a symbol of hands on steering wheel (shown in Figure 7). Post-test surveys queried participants’ preferences on the interfaces employed. The results highlighted some differences with what the literature had suggested, as we explain next. Additionally, the majority of the participants (77.8%) indicated a preference for additional warning though the use of vibration, either of the seat (53%) or of the steering wheel (24.8%).

Aural Warning

91% of participants reported that they found the aural warning helpful; however, 16.7% found it distracting and one participant said it hindered their ability to take control. 80% of the participants indicated that they preferred a human voice over a beep or a solid tone, which was in accord with previous literature. Contrary to previous literature, the majority (91.2%) of participants indicated that they either preferred a female voice (22.2%),or were neutral on the gender selection (69%).

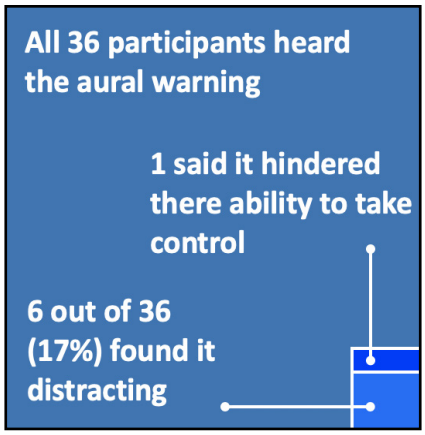

Figure 41. Summary of Main Findings Related to the Aural Warning Employed

The average drift of the 6 participants noted in Figure 41 was 1.36 m, below the average drift of the population (which was 1.45 m). The participant that said it hindered their ability to take control had a maximum drift of 1.29 m, again well below the average of the population.

Visual Warning

We asked participants a range of experience and preference questions about the visual warning interface provided in the central console. Figure 42 summarizes the findings. The most significant finding to note was that 50% of the participants reported not seeing the visual warning. Out of the participants that did see it, two reported it to be distracting, but that it did not hinder their ability to regain control. The finding that many of the drivers did not see the warning is consistent with what was shown in Figure 36: specifically, the weighted average indicated a low score of 1.58 focus level for the central console area, with the majority of the participants keeping their gaze towards the outside environment. This thus prompts an important recommendation towards the development of different modalities of display for visual warning; for example, head-up displays that get projected on the windshield might be more noticeable. Indeed, many participants expressed preferences for the warning location to be Heads-Up Displays (HUD)/windshield displays (36%) or the dashboard (30%), above other presented options, shown in Figure 43. This shows that the majority of participants prefer a visual warning that is directly in front of them. This again agrees with the gaze levels expressed by participants in Figure 36, where 75% expressed high rankings to the front of the vehicle.

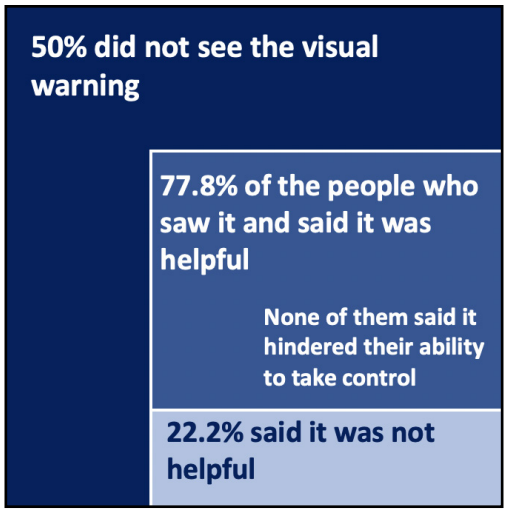

Figure 42. Summary of Findings Related to the Visual Warning

Additional preferences were queried. 64% of the participants indicated red as the preferred color for the visual, instead of the literature suggesting yellow. 64% also stated that they preferred visuals with flashing text (no text was featured in our visual icon). 78% of the participants indicated preference for the wording “Take Control” during the takeover request.

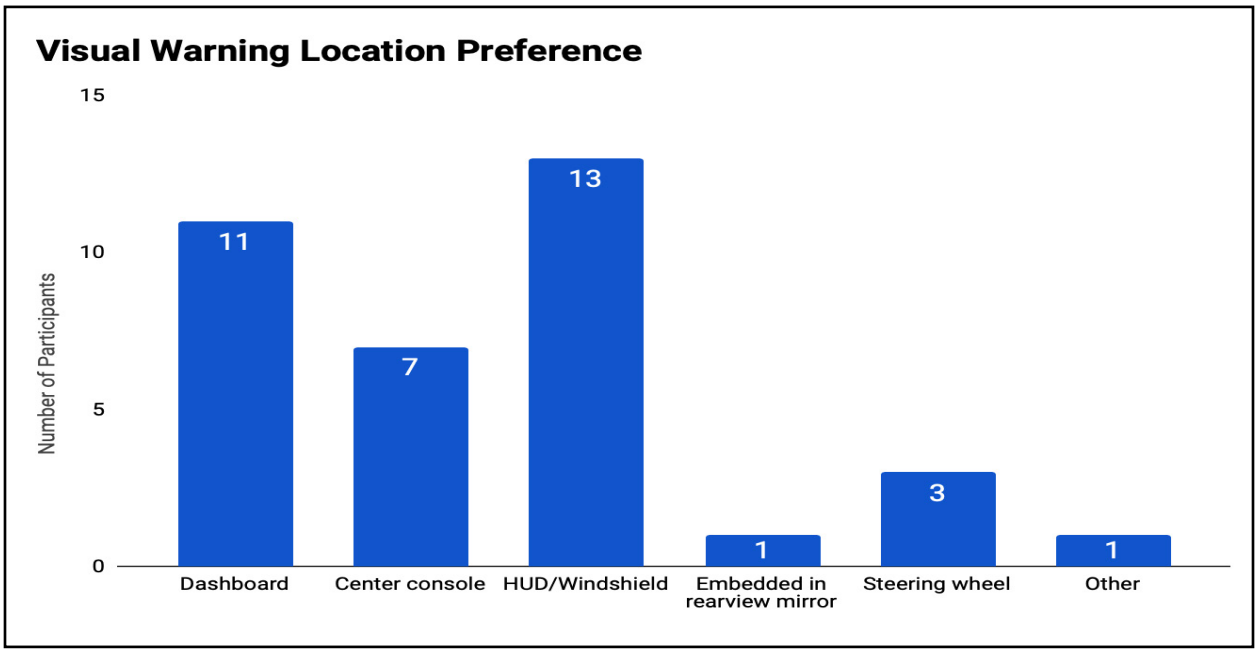

Figure 43. Location of Visual Warning Preferred by Participants

Other Warnings

Finally, participants were asked what other kind of warning they would have liked to receive. Figure 44 shows the different options provided to them. The majority of participants expressed that seat vibration would be their preferred method of being alerted to a disengagement. Only three participants expressed a preference for automatic braking, despite this being a common option in current vehicles currently deployed in the market.

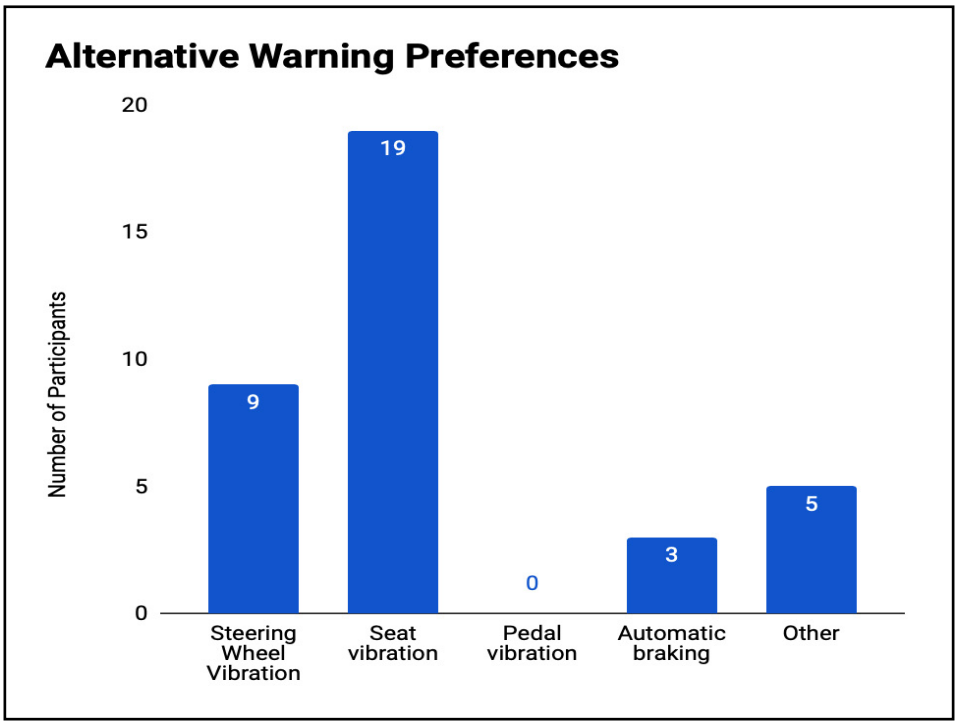

Figure 44. Alternative Warnings Options and Expressed Preferences by Participants

V. CONCLUSIONS AND OPERATIONAL IMPLICATIONS

The analysis of the data described in this report shows several interesting results that lead to numerous venues for future research. Note that all the following conclusions are to be considered preliminary in nature given the small sample size investigated.

The analysis started by looking at drivers' response times to three different inputs: steering; throttling; and braking. A notable result was the observation of a one order of magnitude difference between the peak steering and acceleration response time, and the braking response time. Only 22% of the participants resorted to braking rather than throttling as their first pedal use. This was an unexpected result given the specific geometry of the disengagement on an S-curve. Moreover, an interesting observation was that, of the 78% of the population that engaged in acceleration, a third the participants wrongly recollected and assessed that their response to the disengagement had been to brake and slow down the vehicle. Steering was found to always be the first recorded input, thus prompting us to recommend that the "time to first input" analysis should be avoided, and prefer to it a separate discussion of steering, acceleration, and braking. The factors that were found to affect steering response time were the speed setting (high vs. low) and the age of the driver. For low-speed settings, older drivers achieved a better performance (lower response time) than younger participants; for high-speed settings, the performance was similar for all age groups.

A detailed analysis of drift was executed in order to quantify the quality of the control takeover, and in order to assess the possibility of unintentional lane departures. This analysis also led to the possibility of investigating the participants' perception of success in the recovery (in the presence of traffic), in the form of a binary variable of being capable of, or incapable of, remaining entirely within the assigned lane of travel following the disengagement. Moreover, we compared drivers' drift performance during takeover to their drift performance during entirely manual driving, in order to offset as much bias as possible from the simulated driving.

We first considered the effect of speed on drift, with a low-speed class (55 mph) and a high-speed class (65 mph). The ANOVA showed the highest effect for this factor for both drift metrics considered (maximum lane offset and integral offset). For both metrics, the low-speed category yielded better performance for all test subjects. In automated mode after disengagement, the average maximum lane offset increased in automated mode after disengagement from 0.92 meters at low speed to 1.99 meters at high speed (116% increase). Similarly, for the automated test, the average integral offset showed an increase of 61% from low speed to high speed, and the average integral offset ratio showed an increase of 56% from low speed to high speed (both percentage being computed after removal of outliers). Additionally, the lower speed setting led to a smaller variability in the performance comparison between the manual training and the automated test. This result confirmed what intuitively expected, given the greater ease of control of the vehicle at lower speeds. From an operational standpoint, these results point to the importance of setting up thresholds for maximum operational speed of vehicles driven in automated driving mode when the human driver serves as back-up, perhaps warranting a lower speed limit than conventional vehicles. For example, Tesla's early versions of "Autopilot" presented a

threshold of 50 mph in highway environments that was later on removed; we suggest that features of this sort may indeed be warranted. Our results suggest that the establishment of an operational threshold could reduce the maximum drift and lead to a better controlled takeover. This is a very important point to consider in light of the recorded unintentional lane departures. While participants were instructed to remain within the same lane of travel, unintentional lane departures were recorded in 69% of the cases. Another alarming result was the perception of success in those situations, with all participants but one claiming success in the endeavor. From an operational standpoint, this preliminary finding can serve as a useful starting point for discussions on possible dedicated lane usage for automated vehicles and separation from conventional traffic, as well as for the possibility of increasing lane width in dedicated lanes for semi-automated vehicles.

The second variable investigated was the duration of the AT engagement. The ANOVA showed a borderline significant effect of duration on maximum lane offset ($p = 0.046$), and no significant effect of duration on the integral offset ratio. Additional and more refined testing might be needed on the effects of this variable, which governs important operation considerations in relation to the maximum allowed time an automated driving system can be engaged for, and in defining thresholds for automatically reverting control to the human driver at predetermined intervals. Overall, the lowest engagement duration of 0 to 10 minutes showed the best results in terms of average maximum lane offset. While it can be expected that distraction of the participants led to worst performance for longer tests (and indeed deterioration was observed for the second duration bucket of 10–20 min), male participants tended to reengage and improve the overall metrics for the longest 20–30 minutes setting. The duration of engagement was also found to affect the perception of success obtained by the participants, with higher accuracy obtained for low duration of engagement (0 to 10 minutes).

The third variable, driver's age, revealed rather unexpected results. The ANOVA showed no statistically significant relationship between drift and age of the driver (marginal significance was found for the integral ratio, with a p value of 0.0509); similarly, no significant relationship was found between age and steering response time. Indeed, for both males and females, the age group of 55+ performed best in terms of both maximum lane offset and offset integral. The average maximum lane offset is 31% and 34% higher for the age groups of 35–55 and 18–35 years of age, respectively, than the age group of 55+. Similarly, the average lane integral offset in automated mode is 10% and 7% higher for the age groups of 30–55 and 16–29 years of age, respectively, than the age group of 55+. The age group of 55+ performed best throughout all time groups. Furthermore, the age group of 55+ performed best in terms of mean maximum lane offset and mean lane integral offset for all time groups. The age group of 55+ also showed the least performance differences between fully manual driving and automated driving with disengagement, both for men and women, and for all different times of disengagement. Given the small sample size, these results, like the rest, should be considered preliminary. If valid, however, a possible explanation could be that older subjects not as trusting and not being as familiar with the system and that they are, therefore, more prone to carefully monitor the vehicle. Overall, neither the response times analysis nor the drift analysis support any claim to limit the maximum age of drivers of semi-autonomous vehicles. Note, however, that the oldest age of a participant was 65 and that all drivers were medically fit to drive.

Finally, we investigated situational awareness, emotional and physical response, and preferences related to the human machine interfaces employed. Notable results in these regards include a low accuracy in recalling and estimating the speed of the vehicle, as well as a tendency to overestimate the duration of engagement of the technology. Higher speeds led to more pronounced changes in the level of trust (between, before, and after the test) in the technology as well as higher reported nervousness and fear. Low-speed settings showed higher levels of overall trust. A remarkable observation with important operational implications was that 50% of the population tested did not see the visual warning provided in the central console 10.2-inch display. When asked, 76% of the population expressed a preference for HUD displays, which are just now making their way on the market.

Wherever possible, the results were traced back to notable literature on the topic and were found to be in good accordance. Future work will include further investigations of the effects of time of engagement, as well as validation of the results for larger populations.

APPENDIX

PRE-TEST QUESTIONNAIRE

Participant ID number: Date of test:

Age: Time of test:

Sex:

Driver License History:

1. In which state do you hold your current driving license?

2. Have you acquired a driver's license in any states other than the one mentioned above? If yes, please list all states you have held a driver's' license in. Y/N

3. Have you acquired a driver's license in any country other than the one you currently hold? If yes, please list all the countries you have held a driver's' license in. Y/N

4. When did you acquire your most recent driver's license?

5. Did you go through any form of driving training from an institution before receiving your driving license? If yes, please write the institution(s) you received the training from and the duration. (e.g. 20 hours) Y/N

6. How many times have you taken a written exam to acquire your current driver's license?

7. How many times have you taken a practical (road) test to acquire your current driver's license?

Please leave the following questions blank if you answered No to both 2 and 3.

8. How many times have you taken a written exam to acquire your past driver's license(s)? (Include state(s)/countries)

9. How many times have you taken a practical (road) test to acquire your past driver's license(s)? (Include state(s)/countries)

Driving Experience/Accident History:

1. How many years have you been actively driving?_____

2. When was the last time you drove a car?

3. Have you ever been part of an accident/collision? If yes, when was your most recent accident/collision? Y/N

4. If you answered yes to 3, please list the number of times you have been part of a collision/accident.

5. If you answered yes to 3, please list the number of times you believe you were partially responsible for the collision?

6. If you answered yes to 3, please select any other factors that you believe would have been responsible for the collision.
   a. Outside distraction or disturbance
   b. Weather
   c. Road and infrastructure condition
   d. Recklessness of other drivers
   e. Inside Distraction (e.g., infotainment, cell-phone, etc.)

7. What type of car do you drive most?
   a. Compact
   b. Sedan
   c. SUV
   d. Pick-up Truck
   e. RV
   f. Other

8. What is the make and model of the car you drive most frequently?

9. Does the car you drive have any kind of autonomous features?
   a. Adaptive Cruise Control Y/N
   b. Lane-Keep Assist Y/N
   c. Automatic braking Y/N
   d. Automatic Parking Y/N
   e. Sign Recognition Y/N
   f. Steering Assist Y/N
   g. Blind Spot Detection Y/N
   h. Please list any other autonomous features you have

10. If you selected any of the options in question 9, do you use any of the previously indicated features? Y/N. List the ones you use/have used.

11. If you answered yes to 10, how often do you use these autonomous features?
    a. Less than once/week
    b. 2–3 times a week
    c. Daily
    d. Other

Physical/Medical conditions

1. How many hours of sleep did you get last night? ____

2. On a scale of 1 - 5 (1 being the least and 5 being the most), rate the extent to which you are experiencing the following feelings today, with regard to the simulation you are about to go through.
    e. Fatigue ____
    f. Concern for your physical safety ____
    g. Concern for mental wellbeing ____
    h. Anticipation ____
    i. Anxiety ____
    j. Fear ____
    k. Trust ____
    l. Other _____

3. If you are currently employed, briefly describe the duties of your job.

4. Does your job require activities that put physical strain on your body? If yes, briefly describe the kind of physical activity required for your job. Y/N

5. Have you had any medical conditions in the past that prevented you from driving? If yes, please list and describe them.
    a. Impaired vision
    b. Impaired hearing
    c. Cardiovascular diseases
    d. Nervous system diseases
    e. Respiratory diseases
    f. Psychiatric diseases
    g. Effects of Anesthesia/Surgery
    h. Other (please describe)

Additional Questions:

1. On a scale from 1 to 5, with 1 being not likely and 5 being very likely, how likely are you to buy an autonomous vehicle in the future?

2. On a scale from 1 to 5, with 1 being not at all and 5 being very much, how would you rate your trust in this technology?

POST-TEST QUESTIONNAIRE

Participant ID number: Date of test:

Age: Time of test:

Sex:

Situational Awareness

1. Describe your last known location within the simulation environment.

2. Describe your last known heading within the simulation environment.

3. What do you believe was the speed at which the vehicle was traveling?

4. Is you answer to #3 based on an actual reading of the speedometer?

5. Did a disengagement happen? Y/N

6. Briefly explain what you think would have happened after the disengagement if you did not resume driving.

Disengagement

1. Briefly describe what happened when the disengagement occurred.

2. Were you able to take control of the vehicle after the disengagement? Y/N

3. If you replied yes to #2, describe what you think the vehicle response would have been had you not taken control.

4. What were your first instincts at the time of the disengagement?
   a. Reach for the steering wheel/change vehicle heading
   b. Brake
   c. Combination of braking and steering

d. Accelerating
e. Other

5. Do you believe you were successful in recovering from the disengagement? Y/N/ No disengagement happened

6. Assign a ranking from 1 to 5, with 1 being not at all and 5 being greatly, to each of the following options that describe what you were paying attention to the moment just before the disengagement happened.
   a. Outside the vehicle
      i. To the front
      ii. Situation indicated by side view mirrors
      iii. Situation indicated by rear view mirrors
   b. Inside the vehicle
      iv. Dashboard
      v. Steering wheel
      vi. Center console
      vii. Air vents
      viii. Other
   c. Other
      I. Ipad screen
      II. Cell phone
      III. Water bottle
      IV. Other

Warnings

1. Did you hear an aural warning? Y/N

2. If you answered yes to #1, was the aural warning helpful in alerting you to the disengagement?
   a. Not helpful at all
   b. Somewhat helpful
   c. Neutral
   d. Very helpful

3. If you answered yes to #1, was the aural warning distracting? Y/N

4. If you replied yes to #3, did it hinder your ability to take control? Y/N

5. If you answered yes to #1, were you satisfied with the aural warning provided? Y/N

6. If you had a chance to design an aural warning for your autonomous vehicle, what type of aural warning would you prefer?
   a. Human voice (example: “car disengaging” or “resume driving” or “retake control”)
   b. Beeping
   c. Solid tone
   d. Other

7. If you would have preferred the aural warning to be a voice, would a male or female voice be better?
   a. Male
   b. Female
   c. Neutral

8. If you answered yes to #1, would you have preferred if the aural warning was louder or quieter?
   a. Much quieter
   b. Somewhat quieter
   c. Neutral
   d. Somewhat louder
   e. Much louder

9. Did you see a visual warning? Y/N

10. If you answered yes to #9, was the visual warning helpful in alerting you to the disengagement? Y/N

11. If you answered yes to #9, was the visual warning distracting?

12. If you replied yes to #11, did it hinder your ability to take control of the vehicle? Y/N

13. Were you satisfied with the type of visual warning provided? Y/N/Not applicable

14. What type of visual warning would you have preferred?
   a. Flashing light
   b. Solid light
   c. Flashing text: “Disengagement” (or some other warning message)
   d. Solid text: “Disengagement” (or some other warning message)
   e. Other

15. What color would you prefer the visual warning to be?
    a. Black
    b. White
    c. Brown
    d. Blue
    e. Indigo
    f. Violet
    g. Red
    h. Yellow
    i. Orange
    j. Green

16. Where would you have preferred the visual warning to be located?
    a. Dashboard
    b. Center console
    c. HUD – headboard display
    d. Embedded in rearview mirror
    e. Steering wheel
    f. Other

17. What type of warning message would ~~to~~ you prefer?
    a. “Warning”
    b. “Disengagement”
    c. “Resume Driving”
    d. “Take Control”
    e. Other

18. What other types of warnings would you have liked to alert you to a disengagement?
    a. Steering wheel vibration
    b. Seat vibration
    c. Pedal vibration
    d. Automatic braking
    e. Other

Emotions and Physical state

1. On a scale of 1 to 5, with 1 being not at all and 5 being the most, rank the extent to which you experienced the following feelings during the simulation (assign a number to each of them):
   a. Fear
   b. Anger
   c. Sadness
   d. Surprise
   e. Trust
   f. Anticipation
   g. Adrenaline rush
   h. Nervousness
   i. Other

2. On a scale of 1 to 5, with 1 being not at all and 5 being the most, rate your level of trust in the autonomous vehicle up until the disengagement. ___

3. Did the disengagement experience change your level of trust in autonomous vehicles?
   a. Yes, decreased trust
   b. Yes, increased trust
   c. No, trust level didn't change

4. On a scale of 1 to 5, with 1 being not at all and 5 being the most, rate how much you felt nauseous or motion sick? __

5. If you felt nauseous, did it affect your ability to pay attention and monitor the vehicle?
   a. Not at all
   b. Somewhat
   c. It did significantly

(Questions 5 and 6 were changed of order, used to be 6 first then 5)

6. On a scale of 1 to 5, with 1 being not at all and 5 being the most, rate how much you felt fatigued?__

7. On a scale of 1–5, with 1 being not at all and 5 being the most, how comfortable were you in the simulator. __

8. Do you think this experiment changed your likelihood of buying a semi-autonomous vehicle in the future?
   a. Yes, it increased that likelihood
   b. Yes, it decreased that likelihood
   c. No, it did not affect the likelihood
9. How long do you think you were in the simulator for?
10. How long do you think you manually drove the vehicle?
11. At this time, is there anything else you want us to know?

ABBREVIATIONS AND ACRONYMS

ANOVA	Analysis of Variance
AT	Autonomous Technology
AV	Autonomous Vehicle
CA	California
DOT	Department of Transportation
DMV	Department of Motor Vehicles
FHWA	Federal Highway Administration
FKA	Forschungsgesellschaft Kraftfahrwesen mbH Aachen
FN	False Negative
FP	False Positive
HITL	Human in the Loop
HMI	Human Machine Interface
HSD	Honestly Significant Difference
HUD	Heads Up Display
IRB	Institutional Review Board
MTI	Mineta Transportation Institute
NHTSA	National Highway Traffic Safety Administration
PDF	Probability Density Function
RiSA^{2}S	Risk and Safety Assessment of Autonomous Systems
SAE	Society of Automotive Engineers
SJSU	San Jose State University
SV	Silicon Valley
TN	True Negative
TOR	Take Over Request
TP	True Positive
US	United States
VTD	Virtual Test Drive

BIBLIOGRAPHY

Anderson, J.M., Nidhi, K., Stanley, K.D., Sorensen, P., Samaras, C. and Oluwatola, O.A., 2014. *Autonomous vehicle technology: A guide for policymakers*. Rand Corporation.

Bazilinskyy, P., & De Winter, J. C. F. (2017). Analyzing crowdsourced ratings of speech-based take-over requests for automated driving. *Applied ergonomics*, *64*, 56–64.

Blanco, M., Atwood, J., Vasquez, H. M., Trimble, T. E., Fitchett, V. L., Radlbeck, J., ... & Morgan, J. F. (2015). *Human factors evaluation of level 2 and level 3 automated driving concepts* (No. DOT HS 812 182)

Blommer, M., Curry, R., Swaminathan, R., Tijerina, L., Talamonti, W., & Kochhar, D. (2017). Driver brake vs. steer response to sudden forward collision scenario in manual and automated driving modes. *Transportation research part F: traffic psychology and behaviour*, 45, 93–101.

California Department of Motor Vehicles. (2017). Article 3.7 – Autonomous Vehicles Testing. Title 13, Division 1, par. 227.02 (b).

California Department of Motor Vehicles. (2017). Article 3.8 – Autonomous Vehicles Deployment. Title 13, Division 1, par. 228.06 (a).

Davies A. (2017, January 1). The Very Human Problem Blocking the Path to Self-Driving Cars. *Wired.* Retrieved from: https://www.wired.com/2017/01/human-problem-blocking-path-self-driving-cars/

Dixit, V. V., Chand, S., & Nair, D. J. (2016). Autonomous vehicles: disengagements, accidents and reaction times. *PLoS one*, 11(12), e0168054.

Dogan, E., Rahal, M. C., Deborne, R., Delhomme, P., Kemeny, A., & Perrin, J. (2017). Transition of control in a partially automated vehicle: effects of anticipation and non-driving-related task involvement. *Transportation research part F: traffic psychology and behaviour*, 46, 205–215.

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

Fagnant, D.J. and Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice, 77*, 167–181.

Favarò, F.M., Nader, N., Eurich, S., Tripp, M., Varadaraju, N. (2017). Examining accident reports involving autonomous vehicles in california. *PLoS One*, 12(9), e0184952.

Favarò, F., Eurich, S., & Nader, N. (2018). Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. Accident Analysis & Prevention, 110, 136–148.

Favarò, F. M., Eurich, S., Rizvi, S.S. (2019), *"Human" Problems in Semi-Autonomous Vehicles: Understanding Drivers' Reactions to Off-Nominal Scenarios,* International Journal of Human-Computer Interactions, in print, January 2019.

Forster, Y., Naujoks, F., Neukum, A. and Huestegge, L. (2017). Driver compliance to takeover requests with different auditory outputs in conditional automation. *Accident Analysis & Prevention, 109*, 18–28.

Gao, P., Hensley, R., and Zielke, A. (2014, October). A road map to the future for the auto industry. *McKinsey Quarterly, Oct*.

Gold, C., Körber, M., Lechner, D. and Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. *Human factors*, 58(4), 642–652.

Melcher, V., Rauh, S., Diederichs, F., Widlroither, H. and Bauer, W. (2015). Takeover requests for automated driving. *Procedia Manufacturing*, 3, 2867–2873.

Mok, B.K.J., Johns, M., Lee, K.J., Ive, H.P., Miller, D. and Ju, W. (2015). Timing of unstructured transitions of control in automated driving. *In Intelligent Vehicles Symposium (IV)*, *2015 IEEE* (pp. 1167–1172). IEEE.

Garcia, R., P.E. (2014). Roadway Design Manual. Washington D.C.: U.S. Department of Transportation. Retrieved March 12, 2018, from http://onlinemanuals.txdot.gov/txdotmanuals/rdw/manual_notice.htm

Naujoks, F., Purucker, C., Wiedemann, K., Neukum, A., Wolter, S., & Steiger, R. (2017). Driving performance at lateral system limits during partially automated driving. Accident Analysis & Prevention, 108, 147–162.

NCSL – National Conference of State Legislatures, database on Autonomous Vehicles Legislation. Retreived: from http://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx.

Payre, W., Cestac, J. and Delhomme, P. (2016). Fully automated driving: impact of trust and practice on manual control recovery. *Human factors, 58(2)*, 229–241.

Petermeijer, S., Bazilinskyy, P., Bengler, K. and de Winter, J. (2017). Takeover again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics, 62*, 204–215.

SAE On-Road Automated Vehicle Standards Committee. (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE International.*

Shen, S. and Neyens, D.M. (2017). Assessing drivers' response during automated driver support system failures with non-driving tasks. *Journal of safety research, 61*, 149–155.

Singh, S. (2015). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey* (NO. DOT HS 812 506).

Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., & Hume, R. D. (1979). Tri-level study of the causes of traffic accidents. executive summary (No. DOTHS034353579TAC (5) Final Rpt).

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

Wickens, C. D., Gordon, S. E., Liu, Y., & Lee, J. (1998). An Introduction to Human Factors Engineering.

Zeeb, K., Buchner, A. and Schrauf, M. (2015). What determines the takeover time? An integrated model approach of driver takeover after automated driving. *Accident Analysis & Prevention, 78*, 212–221.

Zeeb, K., Buchner, A. and Schrauf, M. (2016). Is takeover time all that matters? The impact of visual-cognitive load on driver takeover quality after conditionally automated driving. *Accident Analysis & Prevention, 92,* 230–239.

Zhu, W., Zeng, N., & Wang, N. (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.

ENDNOTES

1. Human In The Loop refers to a simulation environment in which a human participant is an active part of the simulation. This implies that the software alone (run through the simulator computer) cannot provide results on its own; it needs to receive specific inputs from a human participant. Typical cases are those used in aviation for pilots, and in the automotive industry for cars. In a way, they resemble more of a “gaming” type of setting, and are particularly useful when researchers need to assess the interaction between a human and a machine element.

2. The assumption behind sample-based simulation is that the population will capture traits that are common to those of typical US drivers. When in a real autonomous vehicle, we do expect that some will rest their hands (albeit temporarily) on the steering wheel. In one situation, a participant had his hands on the steering wheel for the majority of the test. The results for this test were still in the norm and did not show any significant improvement over the others.

3. The maximum number of vehicles present in the simulation at each point in time was limited by the graphical capability of the simulator. Out of the entire vehicle fleet that the simulator could handle smoothly, we wanted to majority of the traffic to travel within the direction of interest for the test. Still, to resemble a real highway, we also wanted traffic to be present in the opposite direction. This led to the choice of the 60/40 ratio. Note that the authors were unable to find definitive literature on the topic, so we relayed on FKA to consult on this, given their extensive experience in highway simulations.

4. Note that in the 2018 revision the DMV removed the requirement to report response times altogether.

5. Note that the ratio is a comparison metric for consistent behavior of a driver (in terms of drift). This implies that two ratios with the same value cannot tell us which driver is doing better in minimizing drift. For example, two participants with a ratio of 1 could be displaying extremely different drift performance, but be equally good (or bad) in both their manual/conventional driving as well as performance after recovery.

6. Tukey’s test showed significant change in the mean only between the youngest age group and the oldest, so these two are the only two groups displayed there.

7. Tukey’s test showed significant change in the mean only between the youngest age group and the oldest, so these two are the only two groups displayed there.

ABOUT THE AUTHORS

FRANCESCA M. FAVARO

Dr. Francesca Favaro is an Assistant Professor in the Department of Aviation and Technology in the College of Engineering at San Jose State University. Prior to joining SJSU she earned a PhD and MS in Aerospace Engineering at the Georgia Institute of Technology, and MS and BS in Space Engineering at Politecnico di Milano, Italy. Dr. Favaro 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 founded the RiSA2S lab, which deals with Risk and Safety Assessment of Autonomous Systems such as drones and self-driving cars. In 2017 she became a research associate of MTI 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 US public roads as well as the National Airspace. She particularly focuses on bridging the gap between the technology world and the current regulatory environment. Dr. Favaro has authored several journal publications and conference proceedings on a variety of topics, and is currently working on a human-in-the-loop study to quantify response times and drivers' reactions to disengagements of the autonomous technology for advanced autonomous vehicles. She has been interviewed by multiple media outlets including the Wall Street Journal, Wired Magazine, and Verge Tech forum as a leading expert in the field of automation safety. Dr. Favaro is an FAA Aviation certified Advanced Instructor, a certified Remote Pilot for drone commercial operations, and a solo-endorsed pilot.

Sky Eurich, Syeda Rizvi, Shivangi Agarwal, Sumaid Mahmood, and Nazanin Nader are all students at San Jose State University.

PEER REVIEW

San José State University, of the California State University system, and the 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.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors

Research Associates Policy Oversight Committee

MINETA TRANSPORTATION INSTITUTE