FOUNDATION FOR TRAFFIC SAFETY

SAFETY RESEARCH USING SIMULATION
UNIVERSITY TRANSPORTATION CENTER

EMERGING TECHNOLOGIES
TECHNICAL REPORT

Human-Machine Interfaces and Vehicle Automation: The Effect of HMI Design on Driver Performance and Behavior

July 2023

Meng Wang, Jah'inaya Parker, Nicholas Wong, Shashank Mehrotra, Shannon C. Roberts
University of Massachusetts–Amherst

Woon Kim, Alicia Romo, William J. Horrey
AAA Foundation for Traffic Safety

607 14th Street, NW, Suite 701
Washington, DC 20005
202-638-5944
AAAFoundation.org

Title

Human-Machine Interfaces and Vehicle Automation: The Effect of HMI Design on Driver Performance and Behavior

(July 2023)

Authors

Meng Wang, Jah'inaya Parker, Nicholas Wong, Shashank Mehrotra, Shannon C. Roberts

University of Massachusetts–Amherst

Woon Kim, Alicia Romo, William J. Horrey

AAA Foundation for Traffic Safety

Author ORCID Information

Wang: https://orcid.org/0000-0002-3304-0610
Parker: https://orcid.org/0000-0001-7583-5174
Wong: https://orcid.org/0000-0002-1707-9769
Mehrotra: https://orcid.org/0000-0002-6749-3773
Roberts: https://orcid.org/0000-0002-0052-7801
Kim: https://orcid.org/0000-0001-9806-1204
Romo: https://orcid.org/0000-0002-4553-1556
Horrey: https://orcid.org/0000-0002-9533-4411

Foreword

This technical report summarizes a study of different human-machine interface (HMI) configurations in helping drivers to return to manual driving in a timely and effective manner, following periods where automation has been active. It builds on recent efforts to identify key guidelines regarding the design and implementation of HMI in automated vehicles. The report should be of interest to researchers and the automobile industry who are working in the domain of advanced vehicle technology.

This report is a product of a cooperative research program between the AAA Foundation for Traffic Safety and the SAFER-SIM University Transportation Center.

C. Y. David Yang, Ph.D.

President and Executive Director
AAA Foundation for Traffic Safety

Dawn Marshall

SAFER-SIM Director
National Advanced Driving Simulator
The University of Iowa

About the Sponsors

AAA Foundation for Traffic Safety

607 14th Street, NW, Suite 701
Washington, D.C. 20005
202-638-5944

www.aaafoundation.org

Founded in 1947, the AAA Foundation for Traffic Safety in Washington, D.C., is a nonprofit, publicly supported charitable research and education organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, individual members, AAA-affiliated insurance companies, and other organizations or sources.

SAFER-SIM University Transportation Center

Federal Grant No: 69A3551747131

SAFER-SIM is comprised of a multidisciplinary, synergistic team of researchers in human factors, engineering, computer science, and psychology who will use innovative simulation approaches ranging from microsimulation to human-in-the-loop simulation to promote safety. SAFER-SIM sponsors research, outreach activities in STEM areas, and aids workforce development efforts in transportation safety.

The universities comprising SAFER-SIM study how road users, roadway infrastructure, and new vehicle technologies interact and interface with each other using microsimulation and state-of-the-art driving, bicycling, pedestrian simulators.

DISCLAIMER

Table of Contents

Executive Summary

The effectiveness of the human–machine interface (HMI) in a driving automation system during takeover situations is based, in part, on its design. Past research has indicated that modality, specificity, and timing of the HMI have an impact on driver behavior. The objective of this study was to examine the effectiveness of two HMIs, which vary by modality, specificity, and timing, on drivers' takeover time, performance, eye glance behavior, and subjective evaluation. Drivers' behavior was examined in a driving simulator study with different levels of automation, varying traffic conditions, and while completing a non-driving related task. Results indicated that HMI type had a statistically significant effect on velocity and off-road eye glances such that those who were exposed to an HMI that gave multimodal warnings with greater specificity exhibited better performance. There were no effects of HMI on acceleration, lane position, other eye glance metrics (e.g., on road glance duration), trust, or usability. Future work should disentangle HMI design further to determine exactly which aspects of design yield differences in safety critical behavior.

Introduction

Vehicle automation technologies have great potential to support drivers and reduce human errors that result in crashes. Driving automation systems have become more widely available and popular over the past several years. Adaptive cruise control (ACC), classified as Level 1 (L1) automation by the SAE International (SAE, 2021), is available on over 90% of new vehicles (Bartlett, 2021). When ACC is combined with lane keeping assistance (LKA), thereby automating lateral and longitudinal control of the vehicle, it is classified as Level 2 (L2) automation. However, at both levels the driver must still be engaged in the driving task and ready to take over quickly. In the future, Level 3 (L3) systems will be available on production vehicles, where the driver is at times no longer responsible for monitoring the road, further altering the relationship between the driver and their vehicles.

While vehicle automation offers drivers both safety and convenience, automation may also have negative consequences. For example, research has shown that drivers were 50% more likely to engage in secondary tasks when using L2 automation compared to when they drove without the technology engaged (Dunn et al., 2019). The burgeoning popularity of vehicle automation technologies and the issue of driver inattention on the road while using these technologies underscores the importance of warnings and alerts systems. When these systems detect driver inattention or road conditions that the automation cannot handle, they issue alerts to the driver to return their attention to the road and/or to take over control of the vehicle.

Warnings and alerts in takeover situations can improve driving behavior and help drivers become more aware of their surroundings. The manner in which an alert or warning is implemented through a vehicle's human-machine interface (HMI) can vary greatly along many dimensions. For example, with respect to the modality of the alert, some studies have found that drivers have faster reaction times for auditory and visual alerts compared to audio alerts alone (Cortens et al., 2019). Others have found that the combination of visual and auditory alerts reduce distraction and inattention within an L2 automated vehicle (Atwood et al., 2019). However, some studies have found that auditory request to intervene led to better or equivocal takeover behavior compared to visual–auditory combinations (Roche & Brandenburg, 2018; Yoon et al., 2019; Zeeb et al., 2015). Additionally, some studies have found that visual alerts only help when they are perceived by the driver (Yang et al., 2018). Although feedback through HMIs has shown to be favored by and useful to drivers, too much information can be overwhelming and poorly designed HMIs can be misleading. Audio alerts, which can be helpful in providing guidance and information to drivers (He et al., 2021; Kasuga et al., 2020; Petermeijer et al., 2017), have been found to be annoying in some situations or implementations (Nakashima & Crébolder, 2010). Moreover, some have found that the effectiveness of audio alerts vary by level of traffic density (Gaspar et al., 2018) and others have shown that visual alerts are only beneficial when they are actually perceived by the drivers

(Yang et al., 2018). Taken together, past research has suggested that the HMI design, specifically alert modality, should depend on the actual driving situation, such as the vehicle automation level, the driving environment, and the content of the alert.

Alerts that precede the standard warning (i.e., a phased or staged warning) have been shown to improve takeover performance and yield increased scanning to the roadway prior to a takeover event (van der Heiden et al., 2017; Vogelpohl et al., 2018). A directional alert combined with tactile warnings, such as an initial auditory warning that instruct drivers and a seat presenting corresponding haptic information was an effective strategy in grabbing drivers' attention in a non-driving task to warn them about an upcoming takeover (Cohen-Lazry et al., 2019; Fitch et al., 2011).

Vehicle automation level, driving environment, and drivers' behavior are all facets that should be considered when designing effective HMIs. More specifically, when designing an HMI one should consider: (a) the modality: visual, auditory, or a combination; (b) the messaging or content: a short declarative sentence, a general sentence, or use of visual icons; and (c) the timing: when warnings should be issued. As noted, multimodal alerts are generally preferred over unimodal visual or auditory alerts as multimodal alerts are best to capture drivers' peripheral attention when they are not looking at the alert signal (Politis et al., 2017). The specificity of takeover requests also plays an important role in alerting the drivers. Providing an alert such as a 'takeover soon' and 'takeover now' yield better driving performance than messages with lower urgency (Brandenburg & Roche, 2020). A compatible and staged alert, such as an initial auditory alert and yellow visual alerts followed by a takeover request, has been shown to be effective to warn drivers about an upcoming takeover when they were engaged in a non-driving task (Cohen-Lazry et al., 2019; Fitch et al., 2011). In addition, the timing of issuing a warning message is also an influential factor when considering HMI effectiveness: studies have found that 2 seconds may be sufficient time given the proper user interface and alert mode (Mok et al., 2017).

Objective, Hypotheses, and Study Overview

When designing HMIs there are many essential aspects that affect their effectiveness. Mehrotra et al. (2022), identified several HMI design recommendations to alert drivers and efficiently convey warning information, including (a) employing a multimodal HMI, e.g., both the visual and auditory modality; (b) providing clear warning messages that support driver decision-making and response selection; and (c) employing multi-stage alerts to provide drivers with a buffer to process and react to the warnings.

Accordingly, this study aimed to examine driver responses to different HMI configurations, which consisted of auditory and visual warning messages, designed for different automated driving levels and scenarios. More specifically, this study sought to examine the effectiveness of two HMIs, which varied by specificity and timing, on drivers' takeover time and performance in different traffic conditions using different levels of automation. Drivers' eye glance behavior, trust, and perceived usability were also examined.

A driving simulator study was conducted in which two groups of drivers experienced one of two different HMI designs. One group experienced a Staged HMI design, which included a two-stage multimodal alert system, in which the drivers received a visual warning message followed by a non-descriptive auditory beep approximately five seconds later. The second group experienced a Simultaneous HMI design, which included a one-stage multimodal alert system where the drivers received a visual and auditory (i.e., explanatory) message at the same time. Within each group of drivers, the HMI was implemented at four different levels of automation (Level 0 to Level 3), with each encountering an edge-case condition/situation, such as weather conditions, traffic situations, or road conditions. Non-driving related tasks were included during drives using Level 3 automation to simulate possible or likely real-world behaviors.

Based on previous research, it was hypothesized that the Simultaneous HMI would yield faster reactions (e.g., faster takeover times) and superior control performance than the Staged HMI. Second, it was hypothesized that the Simultaneous HMI would result in fewer and shorter off-road glances. However, no differences were expected between the HMIs in terms of glances to a non-driving related task; no matter the HMI, we expected drivers to make substantially more glances to the center console to complete the task. Lastly, it was hypothesized that the Simultaneous HMI design would result in higher trust and usability scores than the Staged HMI design.

Method

This human subjects experiment was approved by the University of Massachusetts Institutional Review Board (Study #3045) and performed in accordance with the ethical standards as described in the 1964 Declaration of Helsinki.

Participants

A total of 54 participants were recruited for this experiment. They were required to be between the ages of 18 to 40 and hold a valid U.S. driver's license. An overview of the participants' demographic information is shown in Table 1. Participants were recruited from the University of Massachusetts–Amherst campus via physical flyers and email advertisements. At the beginning of the visit, participants were assigned to: (a) one of two treatment conditions, either Staged HMI or Simultaneous HMI; and (b) the order in which they will experience the driving scenarios. The condition and order assignments were based on a Latin Square Design counterbalance method. Participants were remunerated for their participation.

Table 1. Participant Demographics

	HMI	
	Staged	Simultaneous
Total Sample, N	27	27
Average Age (SD)	19.1 (4.0)	20.1 (1.0)
Gender		
Female	3 (11%)	8 (30%)
Male	23 (85%)	19 (70%)
Non-Binary	1 (4%)	0 (0%)
Ethnicity/Race		
Asian	4 (15%)	8 (30%)
Black/African American	1 (4%)	3 (11%)
Caucasian	19 (70%)	15 (56%)
Hispanic/Latino	2 (7%)	1 (4%)
Other	1 (4%)	0 (0%)

Apparatus

Driving Simulator

The driving simulator used in this study was a fixed-based Realtime Technologies Inc. (RTI, Ann Arbor, MI) driving simulator, which collected data at 60 Hz (Figure 1). The simulator consists of a fully equipped 2013 Ford Fusion surrounded by six screens with a 330-degree field of view. The cab featured two dynamic side mirrors and a rear-view mirror that provided the rear view of the scenario to participants. In the car, a fully customizable virtual instrument cluster and center stack were employed. The simulator was surrounded by speakers to simulate environmental noise and engine sounds along with accurate doppler, intensity, and shift.

The simulator system utilized specialized software capable of simulating different levels of vehicle automation (e.g., from Level 0 to Level 5); however, only Level 0 (L0) to Level 3 (L3) were used in the current study. For L0, the driver was expected to be in full control of the vehicle and was responsible for monitoring the traffic environment. For L1, the driver was expected to monitor the traffic environment and to be in full control of all aspects of vehicle control except maintaining speed and headway. For L2 automation, which is a combination of ACC and a lane centering system, the driver did not need to control speed/headway and steering but was required to monitor the traffic environment. Lastly, for L3 automation, the driver did not need to control speed/headway or steering, and also did not need to monitor the traffic environment, unless the system requested their intervention.

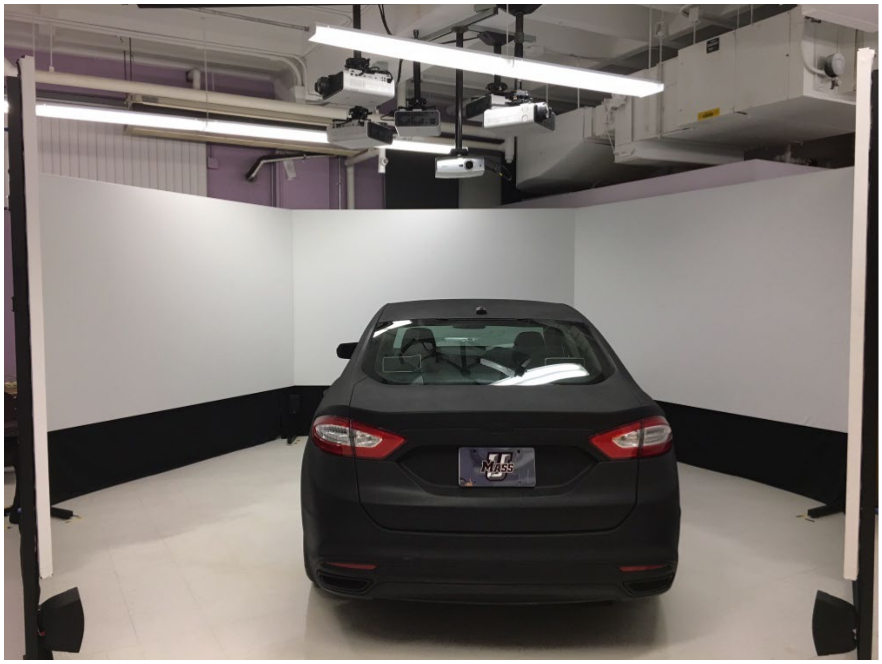


Figure 1. Driving Simulator

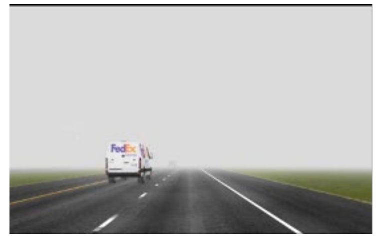
Eye Tracker and Sim Observer

During each drive, participants were recorded using the video capture and review system, Sim Observer (RTI, Ann Arbor, MI). Two cameras recorded the participant’s hand and foot movements, as well as the forward view and instrument cluster. Participants' eye movements were also recorded using a SmartEye tracking system (Gothenburg, Sweden) with three SmartEye Pro cameras mounted on the dashboard and center console. The eye tracking data was used to assess the participants’ attentiveness and focus by classifying their glance regions.

Takeover Scenarios

To assess drivers’ takeover behavior, they encountered four driving scenarios—one for each level of automation from L0 to L3—in which the automation approached the edge of its operational design domain, thereby requiring a transfer of control. For each scenario, the environment consisted of a straight roadway with a speed limit of 65 mph. Each of the four driving scenarios is described below (see Table 2).

Table 2. Driving Takeover Scenarios for Each Level of Automation.

Automation Level	Critical Event	Image
L0 (Manual driving)	Highway driving with foggy weather.	
L1 (ACC)	Highway driving with heavy rain.	
L2 (ACC + LKA)	Highway driving where the lane markings disappear.	
L3 (ACC + LKA)	Highway driving that approaches traffic—a lead vehicle swerving lanes, stop and go traffic, and a police vehicle on the side of the road.	

Foggy weather was introduced during the L0 scenario in order to increase the driving demands as well as the need for drivers to remain attentive. For L1, heavy rain began to fall during the scenario representing a situation where the system should not be used. For L2, the lane markings abruptly disappeared, thereby rendering the lane centering system unusable. Lastly, for L3, the participant encountered a traffic jam wherein a lead vehicle began to behave erratically as other surrounding traffic slowed to a halt. A police vehicle officer was also on the side of the road, creating a plausible cause for the traffic behavior.

Human-Machine Interfaces

Two HMI interfaces were employed: Staged HMI and Simultaneous HMI. The visual and auditory warning messages varied according to the specific HMI. When the automation was engaged and operating as expected, drivers saw the interface shown in Figure 2. For the Staged HMI, the warning messages were the same across all driving scenarios/levels of automation. For the visual warning, there was a red textbox at the bottom of the instrument cluster stating: “Takeover control.” The auditory warning was a simple beep. For the Simultaneous HMI design, the warning messages corresponded to each of the three different scenarios in Table 2. For the visual warning, there was a red textbox in the left-bottom of the instrument cluster and a pictogram at the right-bottom of the instrument cluster. For the auditory warning, a female voice stated a short imperative sentence that corresponded to the visual warning. The warnings from the Simultaneous HMI provided more explicit information to drivers about the appropriate response (e.g., use the brake pedal to disengage ACC). The details of the HMI designs and their visual message, alert signs, and warning message can be seen in Table 3.

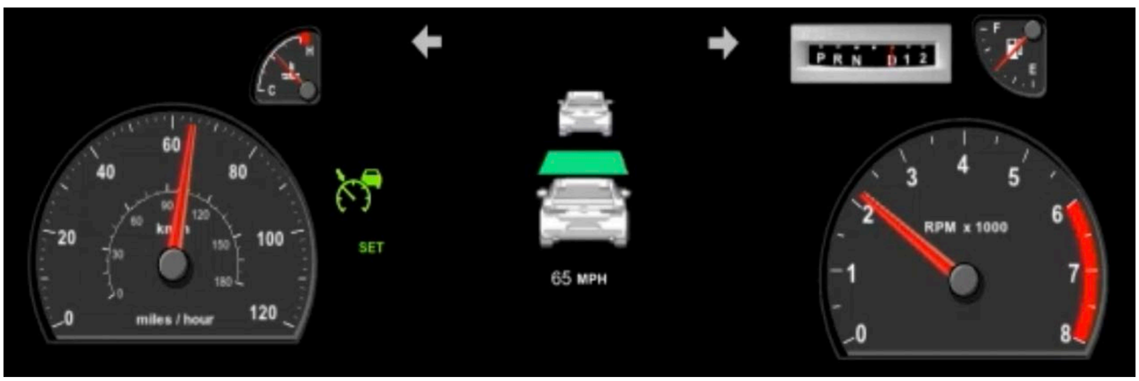
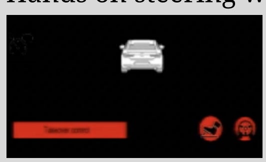


Figure 2. Interface during normal (automated) operations

Table 3. Description of the HMI warnings

HMI	Automation level	Visual warning	Auditory warning
Staged	All levels	Textbox: Takeover control No pictographic icons 	Beep
Simultaneous	L1 (ACC)	Textbox: Takeover from ACC Pictographic: Foot on brake pedal 	“Takeover from ACC”
	L2 (ACC + LKA)	Textbox: Takeover steering Pictographic: Hands on steering wheel 	“Takeover steering”
	L3 (ACC + LKA)	Textbox: Takeover control Pictographic: Foot on brake pedal and Hands on steering wheel 	“Takeover control”

With respect to the timing of the warnings, during each drive the automation disengaged at approximately 2:28 into the drive (see Figure 3). Automation disengaged right before the warnings were issued to emulate situations in which the rise of an edge-case is sudden and unpredictable, thus forcing a disengagement with little or no warning for the driver. For the Staged HMI, the visual warning was issued at approximately 2:30 and remained visible for 2 to 5 seconds; the auditory warning (i.e., a beep) was presented directly afterwards. For the Simultaneous HMI, both the auditory and visual warnings were issued at approximately 2:30. For the Simultaneous HMI, the warnings ended at approximately 2:35, after the auditory alert was completed. A summary of the timing of warnings is shown in Figure 3.

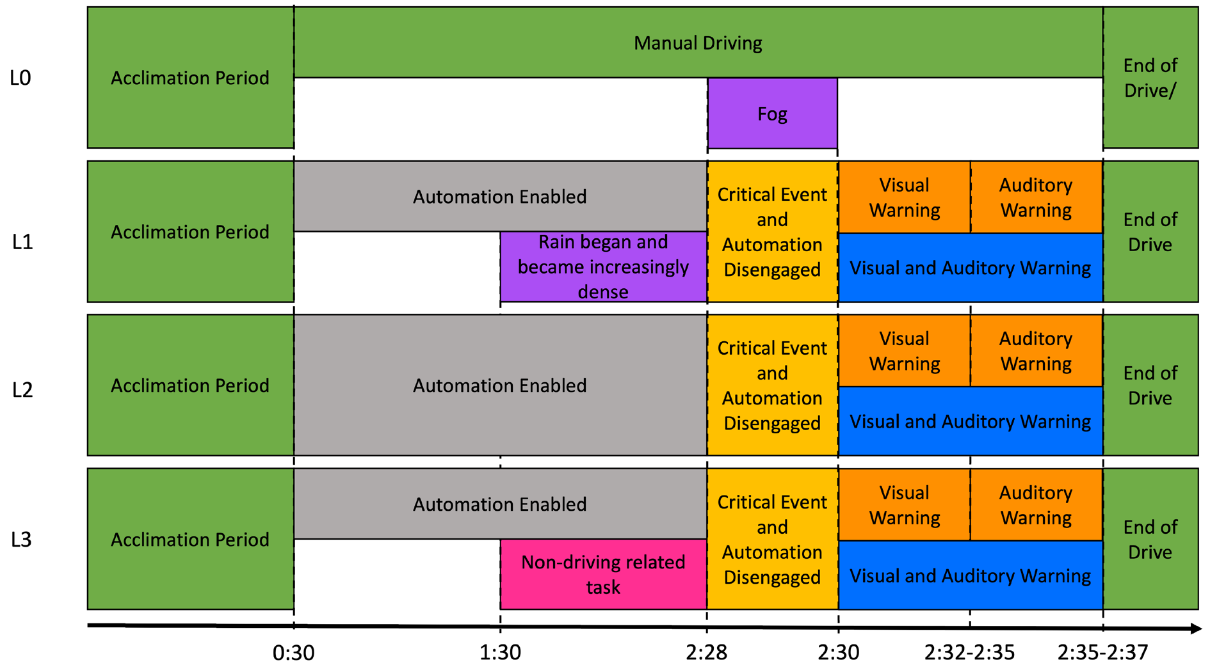


Figure 3. A simulated driving trip showing the sequencing and timing of phases, events, and alerts.

Non-Driving Related Task

During the L3 scenario, the participants were asked to engage in a non-driving related task (NDRT) in order to mimic possible or likely real-world behaviors of drivers using this type of system. A tablet was mounted to the center console of the vehicle (see Figure 4). The participant was instructed to choose from one of three short YouTube videos on the tablet to watch while driving: a cooking video or one of two movie trailers. Once engaged, the NDRT was available before, during, and after the takeover event; however, drivers were alerted by the system when the critical event was encountered (see above). After the drive, participants were asked to complete a form in which they briefly described the YouTube video to ensure that participants were actively engaged in the task during the drive.

Figure 4. An example of the non-driving related task on the tablet

Procedure

When participants arrived at the lab, they were asked to fill out initial screening questionnaires to assure that they qualified for the study. They were then taken through the informed consent process to assure they understood the procedure and what would be required of them. They were introduced to the simulator and the different automated features and capabilities. The researcher proceeded to create a profile in the eye tracking software and complete the calibration process.

The experiment consisted of 5 drives: a practice drive and 4 drives (L0–L3) that emulated a variety of road conditions. For each 3-minute drive, the environment consisted of a straight roadway with a speed limit of 65 mph. Each started with an acclimation period to let the participants familiarize themselves with the driving environment before the automation function was enabled (see Figure 3). In the L1–L3 drives, automation was enabled after the acclimation period. In the L3 drive, once automated mode was engaged for a period, the non-driving related task (NDRT) began. They were told to choose and watch one of the videos once an in-drive audio prompt instructed them to do so. For all drives, the critical event occurred towards the end of the trip and the warnings are issued according to the specific HMI (see Table 3).

Upon completion of the L3 drive, the participant was asked to fill out a short form detailing what happened in the video that they chose. Following completion of the experimental drives, the participants were given questionnaires to assess their perceptions of the HMI they experienced using the System Usability Scale survey (SUS; from Brooke, 1996) and the Trust survey (from Jian et al., 2000), as well as questionnaires to ascertain their basic demographics and driving history. The experiment lasted approximately 60 minutes. Participants were compensated for their time.

Experimental Design

Participants experienced either Staged HMI design or Simultaneous HMI design and drove each of the four automation levels—L0, L1, L2, and L3. The order of automation level was counterbalanced across participants using a Latin Square Design. Though automation level varied, the analyses were conducted separately for each level given significant differences between scenarios and critical events. Given the importance of measuring participants' responses to critical events (i.e., taking over), time (5 seconds before the critical event versus 10 seconds after the critical event) was also considered with respect to driving performance behavior. In summary, the independent variables in this study are:

- HMI design (between subject, 2 levels)
    - Staged
    - Simultaneous
- Automation level (within subject, 4 levels)
    - L0, L1, L2, L3
- Time (within subject, 2 levels)
    - Before takeover (5 seconds)
    - After takeover (10 seconds)

Dependent Variables and Analysis

To measure drivers' performance, takeover time was calculated. Takeover time was defined as the time between when automation disengaged and when the participant began to regain control of the vehicle (i.e., put their hands on the wheel or their foot on the pedal). Before applying the statistical tests, outliers (takeover times greater than 8 seconds) were removed to ensure the data followed a normal distribution. An analysis of variance (ANOVA) was used to analyze takeover time for each level of automation.

Vehicle velocity, acceleration, lane position, and proportion of large deceleration (based on times when deceleration > 0.45 g) around the takeover were also measured. A repeated-measures ANOVA was used to analyze velocity, standard deviation of velocity, mean acceleration, maximum (negative) acceleration, and standard deviation of lane position, after a Box-Cox transformation was applied to ensure the normality of the data distribution. A generalized linear mixed model was applied on the maximum acceleration to examine the effect of HMI types and time. A zero-inflated negative binomial model was applied on the proportions of large deceleration since the data has an excess number of zeros (i.e., no instances of large deceleration).

Eye glance measures were collected during the entire drives; however, mean on-road glance duration, mean off-road glance duration, total off-road glance duration, long off-road glance ratio, and off-road glance frequency were only evaluated during the approximately 15-second windows surrounding the takeover events (approximately 5 seconds before the event and 10 seconds after the event). The actual length of the clip was fine-tuned so that the clip started at the beginning of a glance and ended at the end of a glance, resulting in a slightly longer clip on average than was used for the driving performance analysis. Before applying the statistical tests, any extremely short glances (< 0.16 seconds) or long glances (> 70 seconds), which were indicative of poor eye tracking performance, were removed. The maximum glance duration was 12 seconds after removing said eye glances. A Box-Cox transformation was applied on the mean glance duration and total glance duration to ensure a normal distribution. In the following analysis, two types of glances were considered: (a) off-road glances, which are any glances not towards the windshield, and (b) glances towards the instrument cluster to examine the participants' warning-checking behavior. Repeated-measures ANOVA were applied on the mean glance duration and total glance duration. Zero-inflated generalized linear mixed models were applied on the long-glance ratio and glance frequency towards the instrument cluster. Generalized linear mixed models were applied on the off-road glance frequency.

With respect to the SUS survey and the Trust survey, the aggregated SUS score was calculated based on the methods developed by Brooke (1996) for each participant, and the aggregated Trust score was calculated based on the scale developed by Jian et al. (2000) for each participant. They were analyzed using ANOVAs.

Results

Driving Performance

Descriptive statistics for the driving performance data are presented in Table 4. Takeover time was lower for the Simultaneous HMI group and for L2 automation. Vehicle velocity and acceleration were lower during L3 automation, compared to L1 and L2. The proportion of large deceleration was similar across HMI types and all the levels of automation. A complete detailing of all p-values for the statistical tests is provided in Appendix A.

Table 4. Descriptive statistics of driving performance data; mean values are shown with standard deviation in parentheses

Level of Automation	L1				L2				L3			
HMI Type	**Staged**		**Simultaneous**		**Staged**		**Simultaneous**		**Staged**		**Simultaneous**	
Takeover Time (s)	7.00 (4.55)		4.64 (0.745)		3.25 (1.26)		2.27 (1.28)		5.89 (2.93)		4.9 (2.51)	
	Before takeover	**After takeover**	**Before takeover**	**After takeover**	**Before takeover**	**After takeover**	**Before takeover**	**After takeover**	**Before takeover**	**After takeover**	**Before takeover**	**After takeover**
Velocity (mph)	62.43 (4.06)	60.56 (3.23)	60.07 (6.83)	59.99 (5.88)	64.78 (0.90)	65.66 (1.48)	65.29 (0.78)	66.46 (2.70)	62.47 (2.21)	52.4 (7.15)	63.11 (2.11)	53.9 (6.84)
Acceleration (m/s/s)	-0.21 (0.10)	0.07 (0.29)	-0.36 (0.07)	0.21 (0.15)	-0.25 (0.07)	0.24 (0.08)	-0.17 (0.11)	0.24 (0.20)	-0.40 (0.14)	-0.66 (0.57)	-0.37 (0.12)	-0.69 (0.59)
Maximum (negative) Acceleration (m/s/s)	-0.65 (0.46)	-1.65 (1.89)	-0.48 (0.06)	-0.33 (0.13)	-0.51 (0.02)	-0.25 (0.35)	-0.47 (0.12)	-0.24 (0.22)	-1.13 (1.58)	-4.48 (3.65)	-0.77 (0.70)	-3.51 (2.79)
Proportion of Large Deceleration	0.04 (0.05)	0.05 (0.05)	0.10 (0.19)	0.01 (0.02)	0.01 (0.00)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.05 (0.08)	0.14 (0.12)	0.04 (0.06)	0.16 (0.15)
Standard Deviation of Lane Position	0.08 (0.03)	0.14 (0.04)	0.06 (0.04)	0.15 (0.18)	0.03 (0.03)	0.13 (0.06)	0.07 (0.10)	0.28 (0.36)	0.07 (0.06)	0.26 (0.30)	0.07 (0.05)	0.26 (0.22)

Takeover Time

For each level of automation, mean takeover time for each HMI type was calculated and submitted to a one-way between-subjects ANOVA with HMI type as the independent variable. As shown in Figure 5, takeover time for the Simultaneous HMI was nominally shorter for each level of automation compared to Staged HMI; however, these did not reach statistical significance.

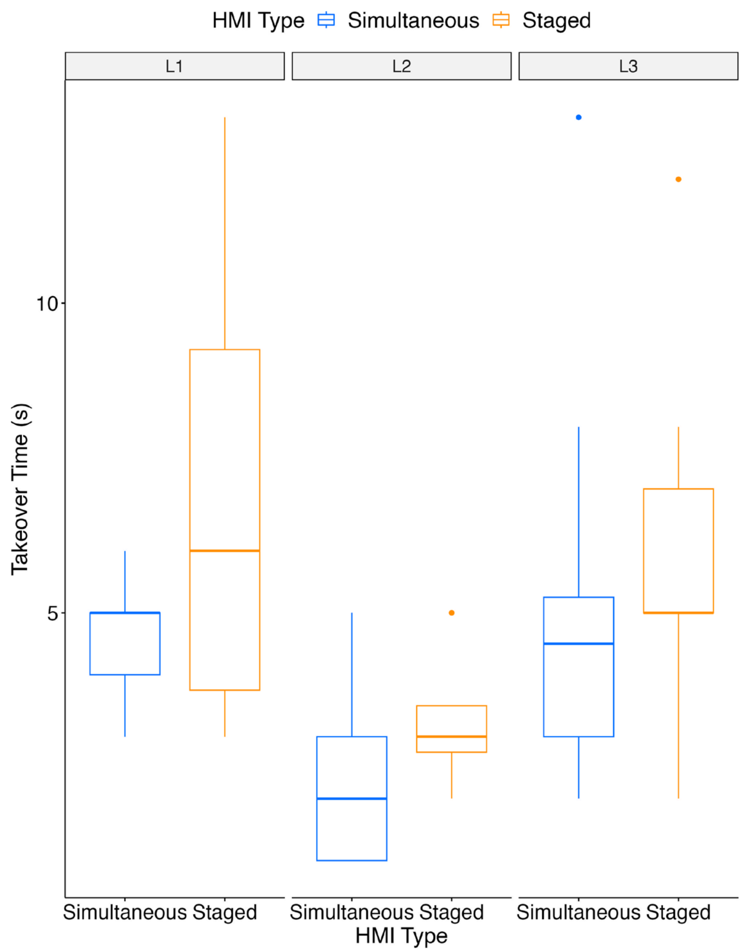


Figure 5: Takeover time by level of automation and HMI

Velocity

For each level of automation, mean velocity before and after automation takeover for each HMI type was calculated and submitted to a two-way repeated-measures ANOVA with time and HMI type as independent variables. The results indicated that for L1 and L2 automation, there were no significant main effects or interactions, whereas for L3 automation, time was significant ($F(1, 27) = 134.8$, $p < .001$). As shown in Figure 6, the vehicle velocity was substantially reduced for both HMI types after takeover under L3 automation.

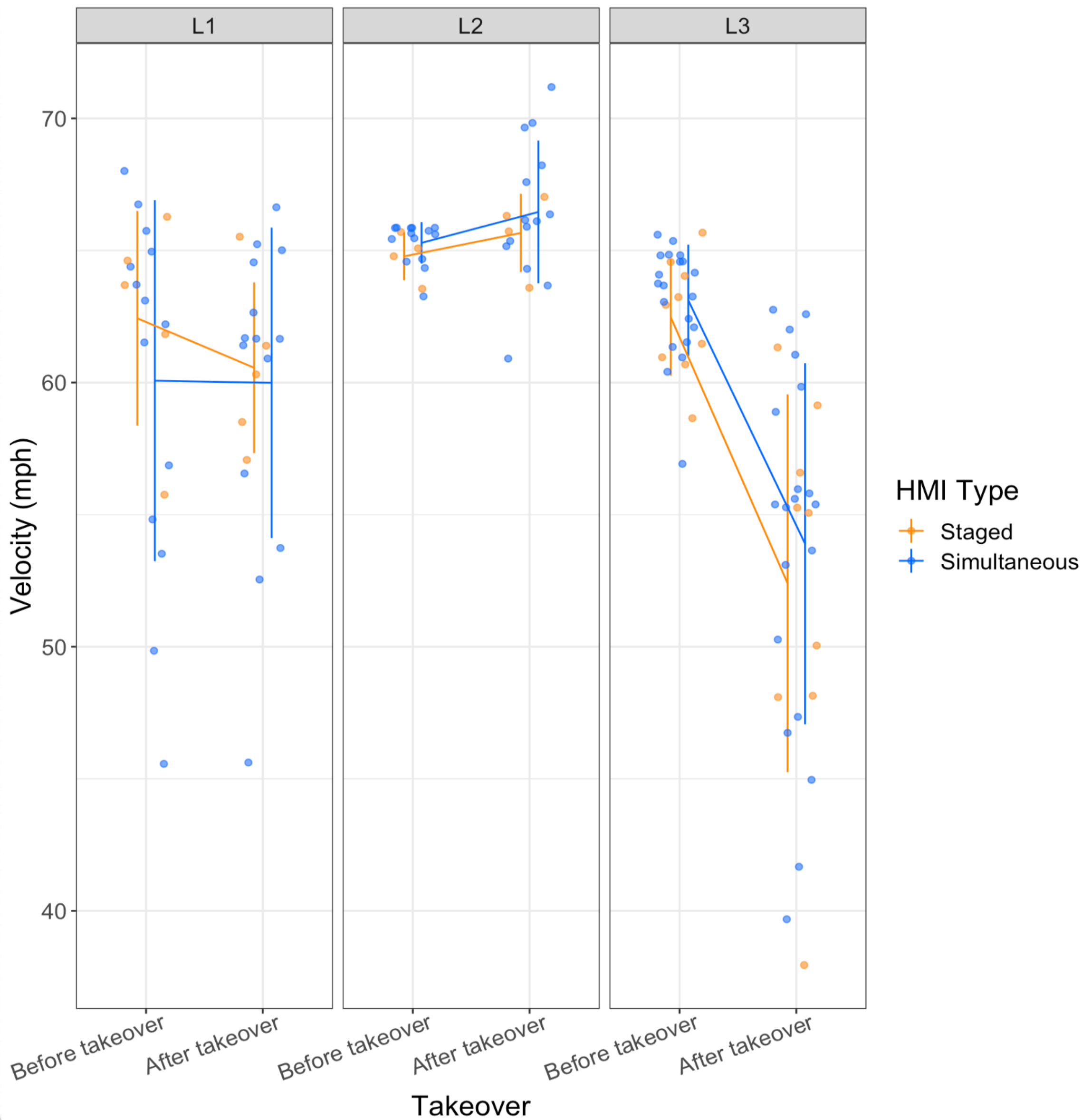


Figure 6: Vehicle velocity by time, level of automation, and HMI type.

Standard Deviation of Velocity

The standard deviation of velocity between before the takeover and after the takeover across different levels of automation was also evaluated. A two-way repeated-measures ANOVA was used to examine the effects of HMI type and time. The results revealed that for L1, there were significant main effect of time and a significant HMI type by time interaction, and for L2 and L3, there was a significant main effect of time. Figure 7 shows that for L1, the standard deviation of vehicle velocity for the Staged HMI increased more after takeover than that of the Simultaneous HMI. For L2 and L3, the standard deviation of vehicle velocity increased for both HMI types.

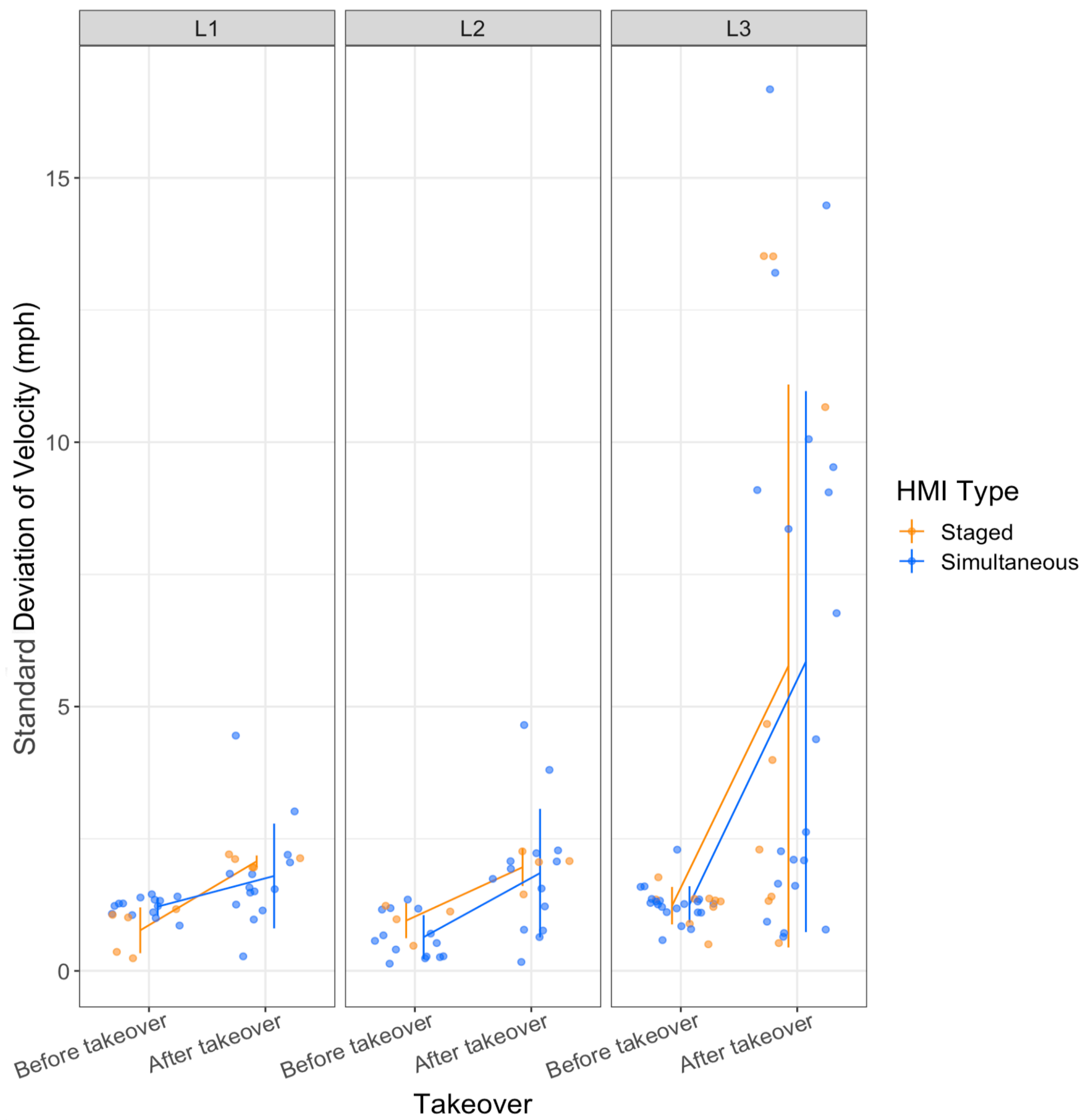


Figure 7: Standard deviation of vehicle velocity by time, level of automation, and HMI type.

Acceleration

For each level of automation, mean acceleration before and after takeover was calculated and submitted to a two-way repeated-measures ANOVA with time and HMI type as independent variables. With L1 automation, there was a significant effect of time ($F(1, 17) = 54.7$, $p < .0001$) and an interaction between HMI type and time ($F(1, 17) = 6.4$, $p < .0001$). When the level of automation was L2 and L3, time was significant (L2: $F(1, 17) = 56.7$, $p < .0001$, L3: $F(1, 17) = 5.7$, $p = 0.02$), but there was no effect of HMI type nor any interactions. As shown in Figure 8, for L1, the increase in acceleration after takeover was greater for the Simultaneous HMI than the Staged HMI. With L1 and L2 automation, there was negative acceleration (i.e., deceleration) prior to the takeover and positive acceleration after takeover. For L3 automation, there was deceleration both before and after takeover.

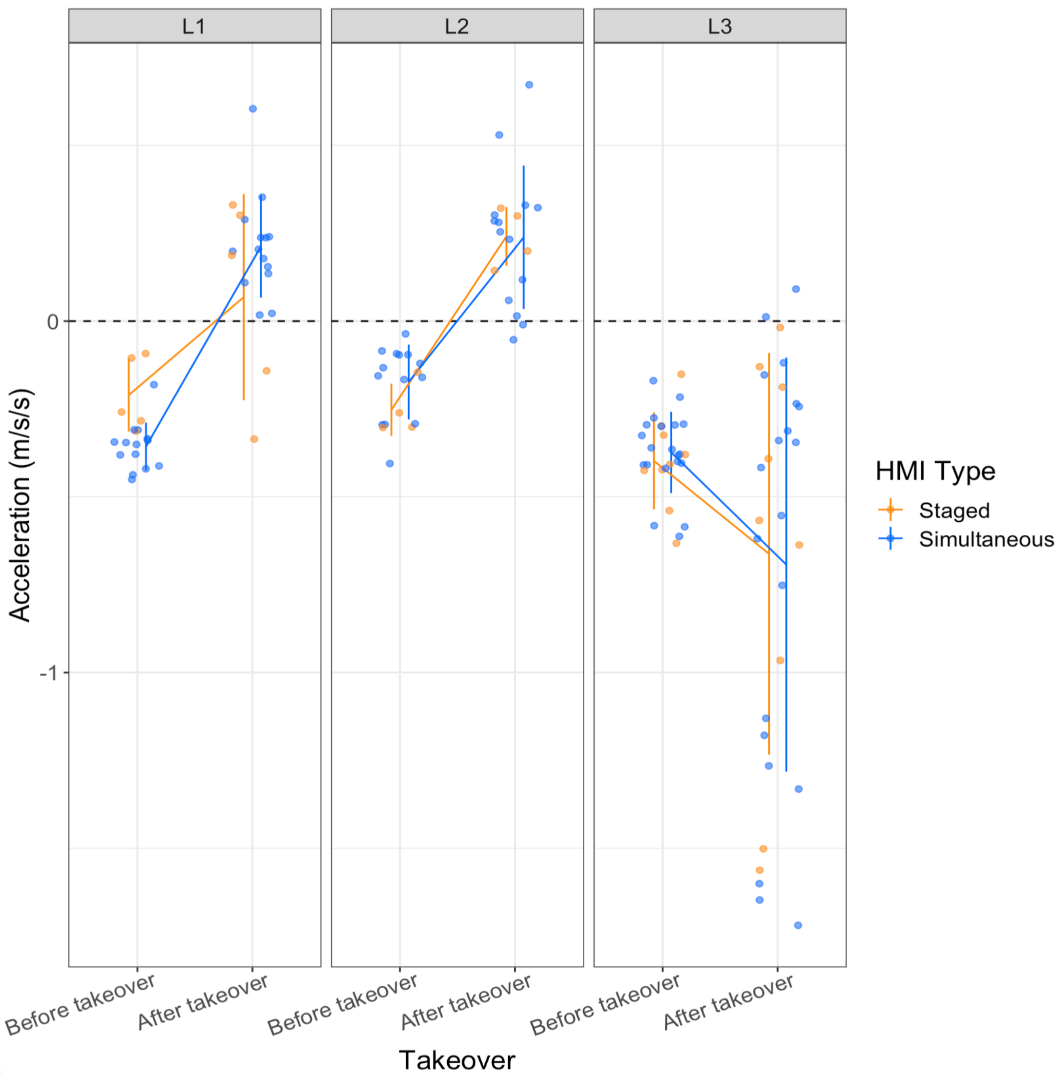


Figure 8: Vehicle acceleration by time, level of automation, and HMI type. The dashed line represents an acceleration of 0.

Maximum (Negative) Acceleration

Maximum (negative) acceleration (i.e., deceleration) was another dependent variable to assess drivers' driving behavior when encountering the critical event, with larger negative values indicating more severe deceleration by the drivers in the scenario. A generalized linear mixed model was used separately on each level of automation with time and HMI type as the independent variables. The results revealed that for L1, time and the interaction of HMI type and time were significant. For L2, there were no significant factors and for L3, time was significant. Figure 9 presents the effects of time and HMI type on maximum (negative) acceleration across different levels of automation.

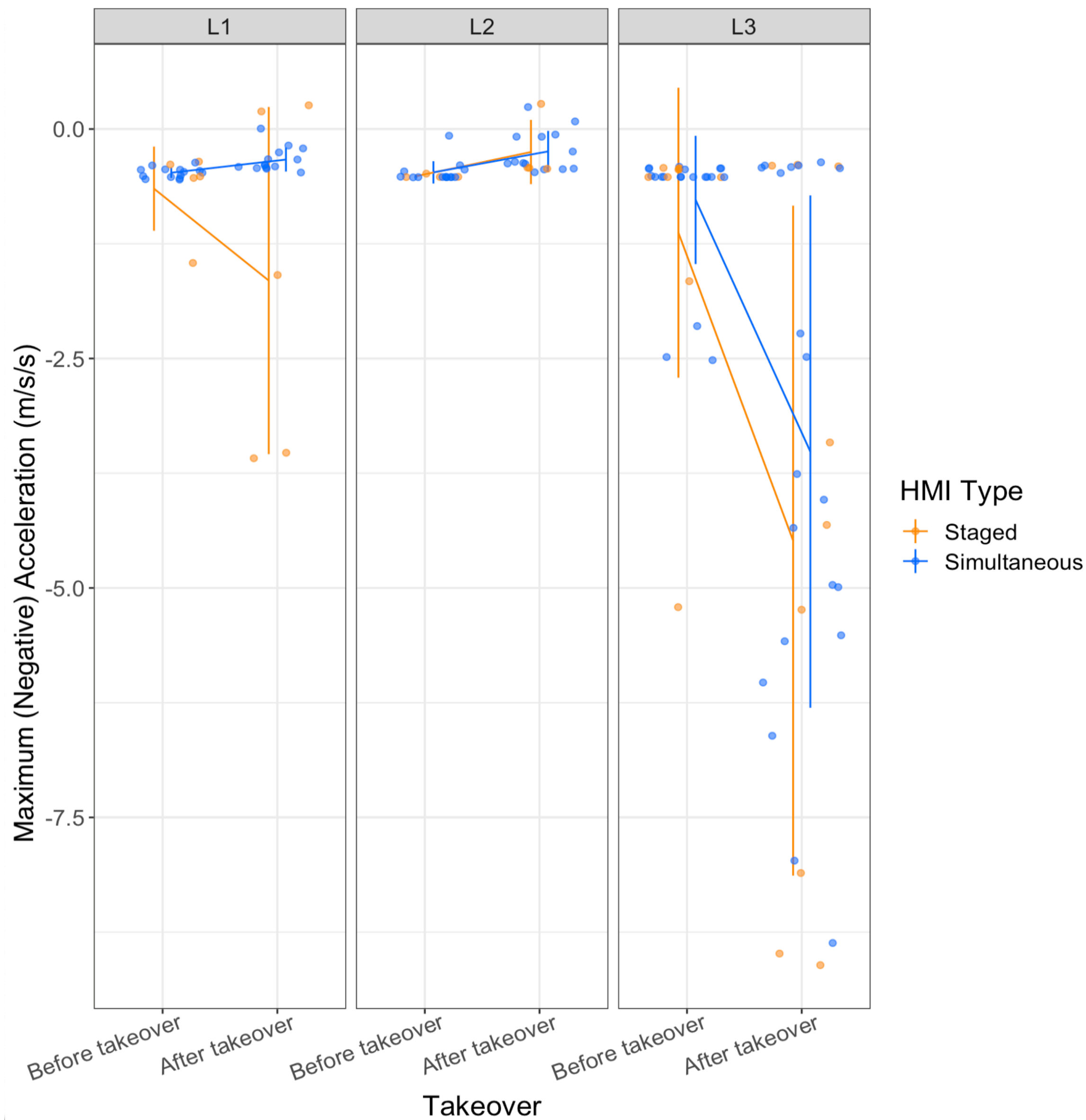


Figure 9: Maximum (negative) acceleration by time, level of automation, and HMI type.

Proportion of Large Deceleration

A zero-inflated generalized linear mixed model was conducted separately on each level of automation with time and HMI type as independent variables. The model results indicated that for L1, the interaction of time and HMI type (beta = -3.2, p = 0.015) was significant. For L2, there were no significant effects, and for L3, time (beta = 1.2, p < .0001) was significant. Figure 10 shows the time proportion of large deceleration across different levels of automation, time, and HMI type.

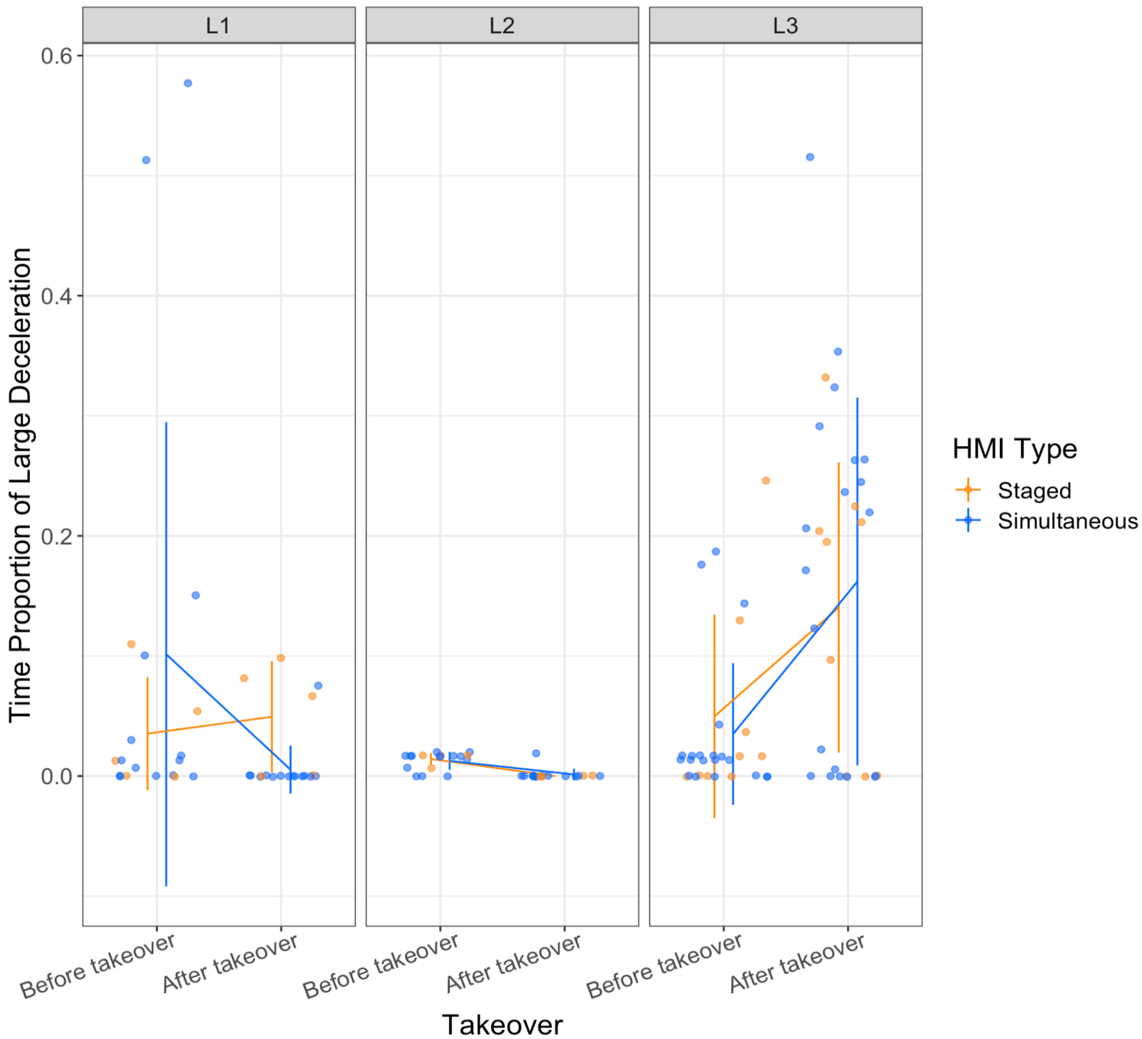


Figure 10: Proportion of large deceleration by time, level of automation, and HMI type.

Standard Deviation of Lane Position

To measure drivers' lane changing behavior, standard deviation of lane position was evaluated using repeated-measures ANOVA. The results and Figure 13 show that only time was significant for L1, L2, and L3.

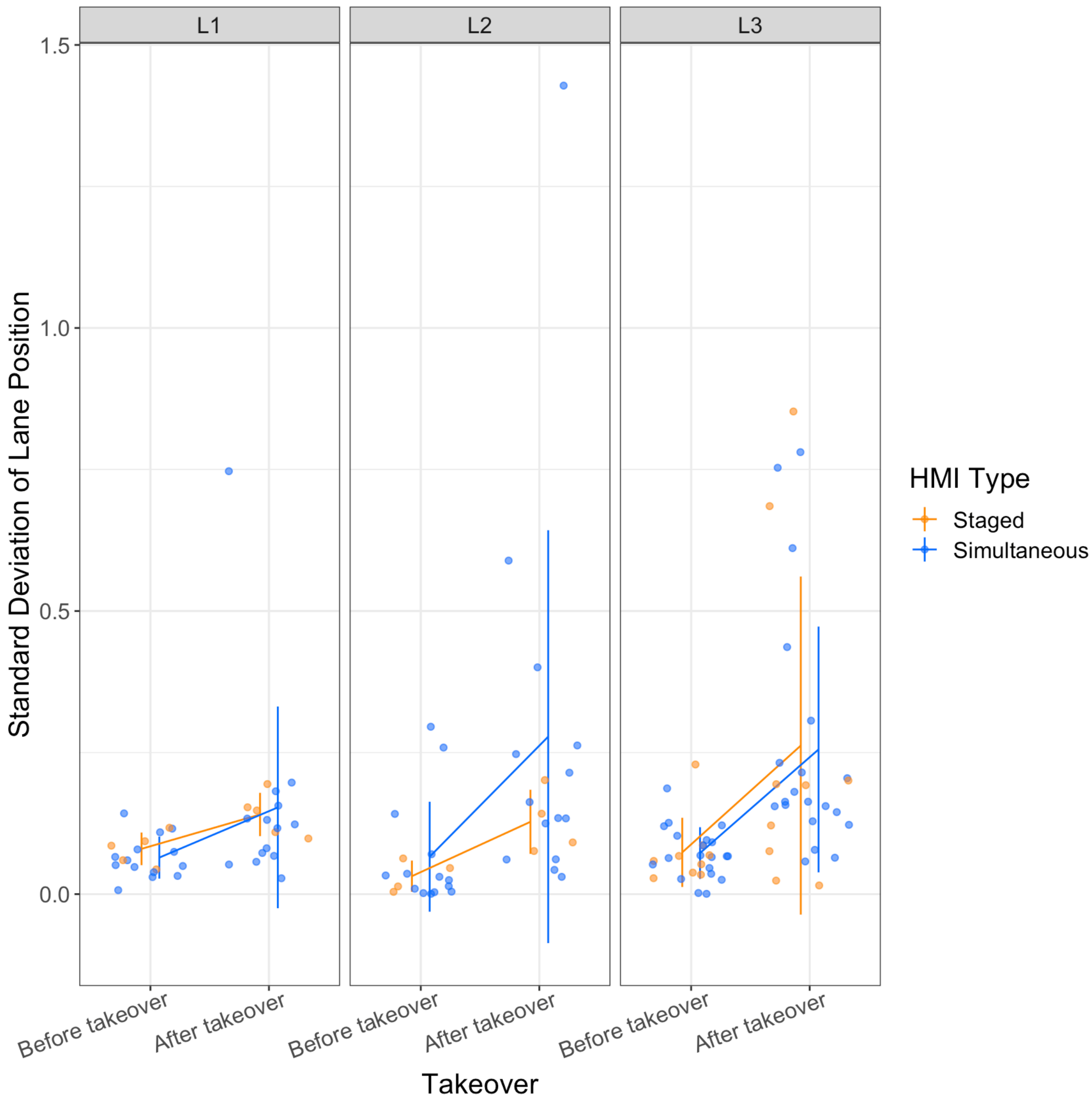


Figure 11: Standard deviation of lane position by time, level of automation, and HMI type.

Eye Glance Behavior

In the dataset, there were six glance areas of interest: center console (i.e., towards the tablet for the NDRT), instrument cluster (e.g., for the speedometer or HMI warning messages), rearview mirror, left screen (i.e., left window), right screen (i.e., right window), and windshield (i.e., forward roadway). The glances towards the windshield were considered on-road glances and glances towards other areas were considered off-road glances. Most glances were towards the center console and windshield and there were more off-road glances than on-road glances.

Table 5 shows descriptive statistics for glances during the approximately 15-second window surrounding the critical event. The data show that the mean glance duration, total glance duration, and long-glance ratio varied across different levels of automation and HMI types. However, the off-road glance frequency and the glance frequency towards the instrument cluster were lower in L3 automation than L1 and L2 automation. A complete detailing of all p-values for the statistical tests is provided in Appendix A.

Table 5. Descriptive statistics for eye glance data during critical events; mean values are shown with standard deviations in parentheses.

Level of Automation	L1		L2		L3	
HMI Type	**Staged**	**Simultaneous**	**Staged**	**Simultaneous**	**Staged**	**Simultaneous**
Mean Glance Duration						
On-road	0.41 (0.33)	0.95 (1.25)	1.89 (4.69)	1.41 (2.19)	1.03 (1.65)	1.07 (1.46)
Off-road	0.52 (0.46)	0.53 (0.36)	0.53 (0.29)	0.84 (1.46)	0.88 (1.51)	0.68 (0.79)
Towards instrument cluster	0.54 (0.26)	0.55 (0.29)	0.44 (0.22)	0.51 (0.39)	0.59 (0.39)	0.57 (0.27)
Total Glance Duration						
Off-road	8.19 (1.85)	6.65 (4.31)	5.27 (3.11)	6.25 (5.77)	7.90 (6.85)	4.96 (5.33)
Towards instrument cluster	3.63 (1.31)	2.46 (1.71)	1.18 (0.8)	1.13 (0.67)	1.88 (1.28)	0.82 (0.64)
Glance Frequency						
Off-road	15.8 (7.05)	12.55 (5.99)	10.00 (4.97)	7.45 (4.41)	9.00 (5.15)	7.31 (5.7)
Towards instrument cluster	6.75 (2.75)	4.50 (1.87)	2.67 (1.53)	2.2 (1.64)	3.20 (2.17)	1.43 (0.79)
Long-Glance Ratio						
Off-road	0.03 (0.06)	0.00 (0.00)	0.00 (0.00)	0.02 (0.03)	0.05 (0.1)	0.04 (0.1)

Glance Duration

There were five dependent variables related to glance duration: mean on-road glance duration, mean glance duration towards the instrument cluster, mean off-road glance duration, total glance duration towards the instrument cluster, and total off-road glance duration. The repeated-measures ANOVA with only HMI type as an independent variable suggested that there were no significant effects for any of these dependent variables. Figure 12 shows the mean off-road glance duration separated by levels of automation and HMI type. The plot indicates that most outliers were glances towards the center console, especially for the L3 automation.

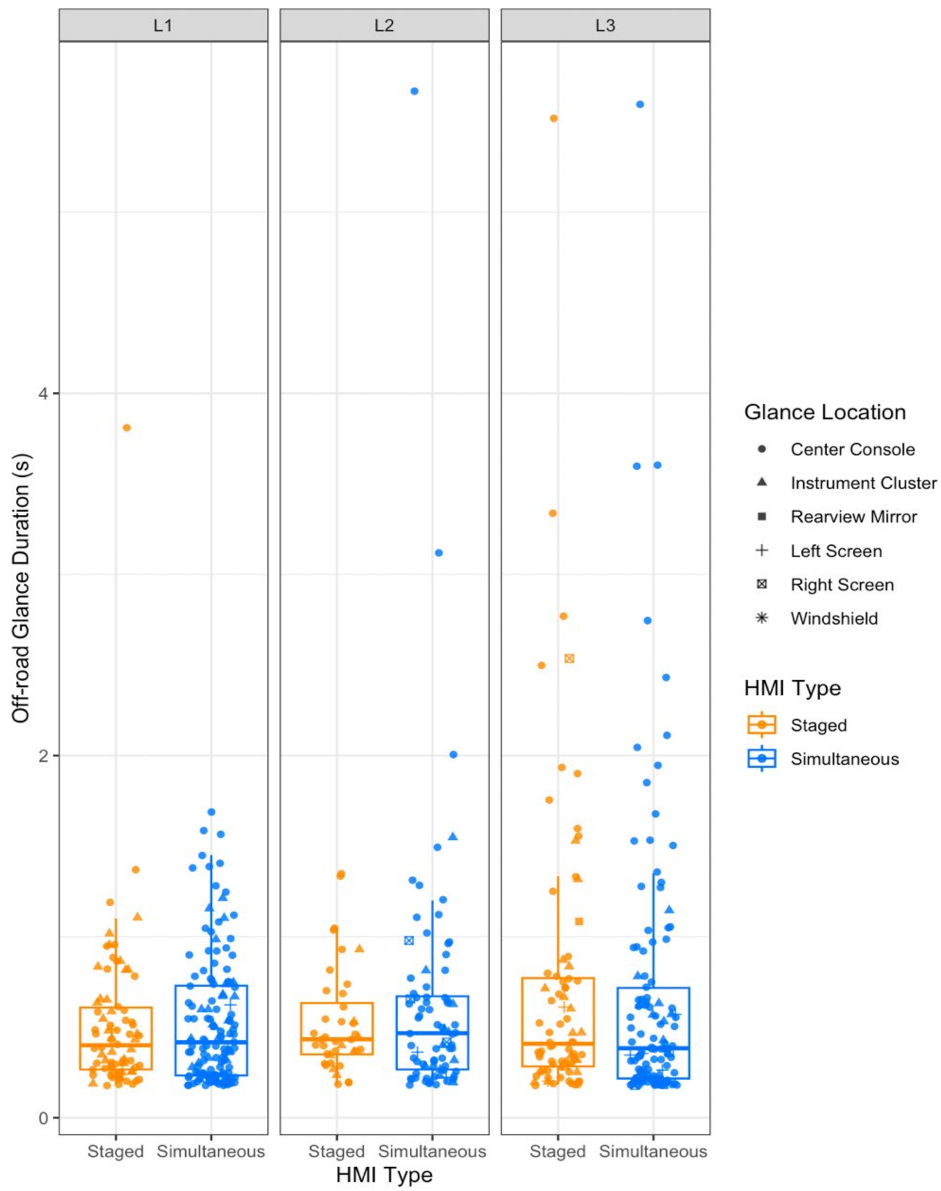


Figure 12: Mean off-road glance duration by level of automation and HMI type; each area of interest is represented by shape.

Long Off-Road Glance Ratio

Only long off-road glance ratio was considered as there were no long glances towards the instrument cluster. The long off-road glance ratio was calculated by the ratio of the number of long off-road glances (> 2 s) to the number of all off-road glances. The data was submitted to a zero-inflated generalized linear mixed model with the HMI type being the only independent variable separated by the level of automation. The zero-inflated generalized linear mixed model results indicated that there was no significant effect of HMI type. Figure 13 shows the long off-road glance ratios, where it can be seen that the majority of values were close to 0. Not surprisingly, these ratios increased for the L3 condition, where the NDRT likely incurred more long duration glances relative to L1 and L2.

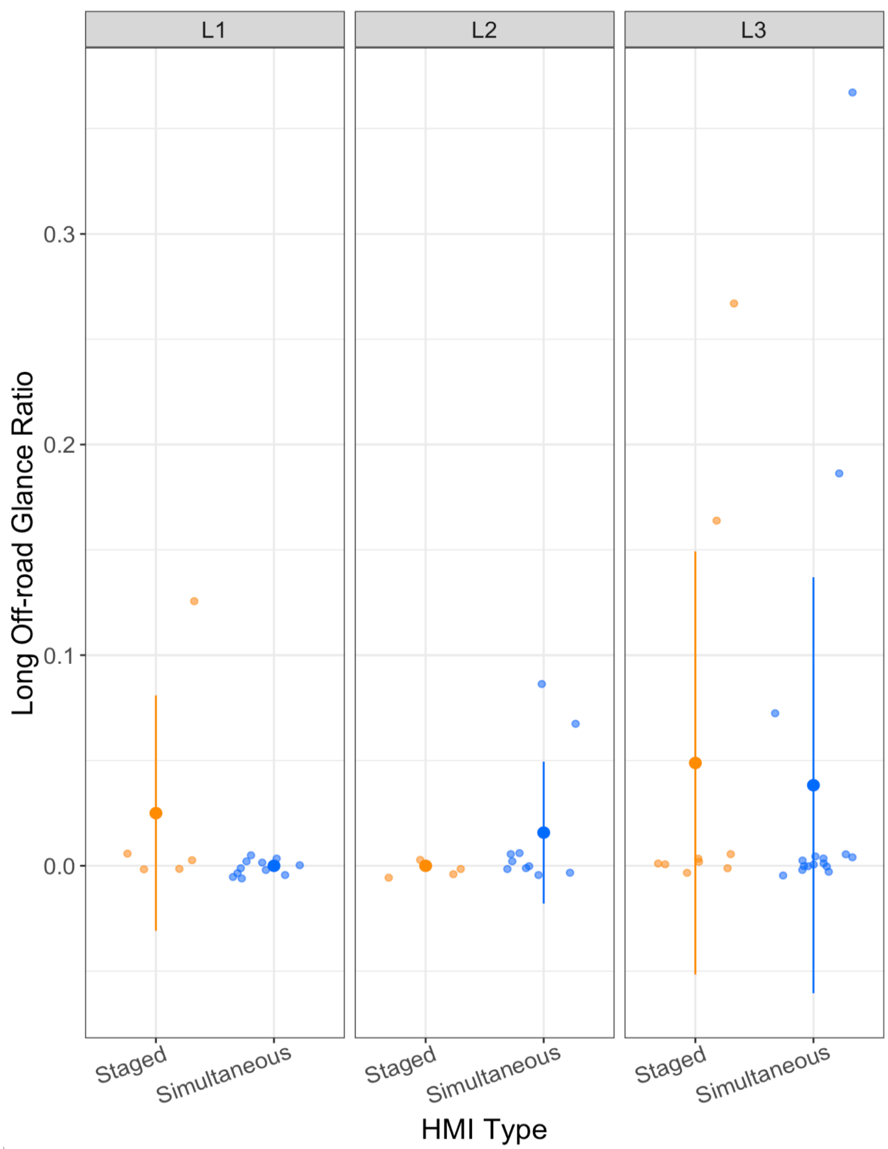


Figure 13: Long off-road glance ratio by level of automation and HMI type.

Glance Frequency

The glance frequency towards the instrument cluster was submitted to a zero-inflated GLMM separated by the level of automation with only HMI type as an independent variable. For L1 and L2 automation, there were no significant effects. When the level of automation was L3, HMI type (beta = 0.58, p = 0.02) was significant.

The off-road glance frequency data was submitted to a GLMM without the zero inflation settings with only HMI type as an independent variable. The results were similar to that of the glance frequency towards the instrument cluster: with L3 automation, HMI type (beta = 0.11, p < .0001) was significant, whereas there were no significant effects for L1 or L2 automation.

Figure 14 shows boxplots of glance frequency separated by levels of automation, HMI type, and glance location. It can be seen that the glance frequency was reduced for the Simultaneous HMI type under the L3 automation for both the off-road glances and glances towards the instrument cluster.

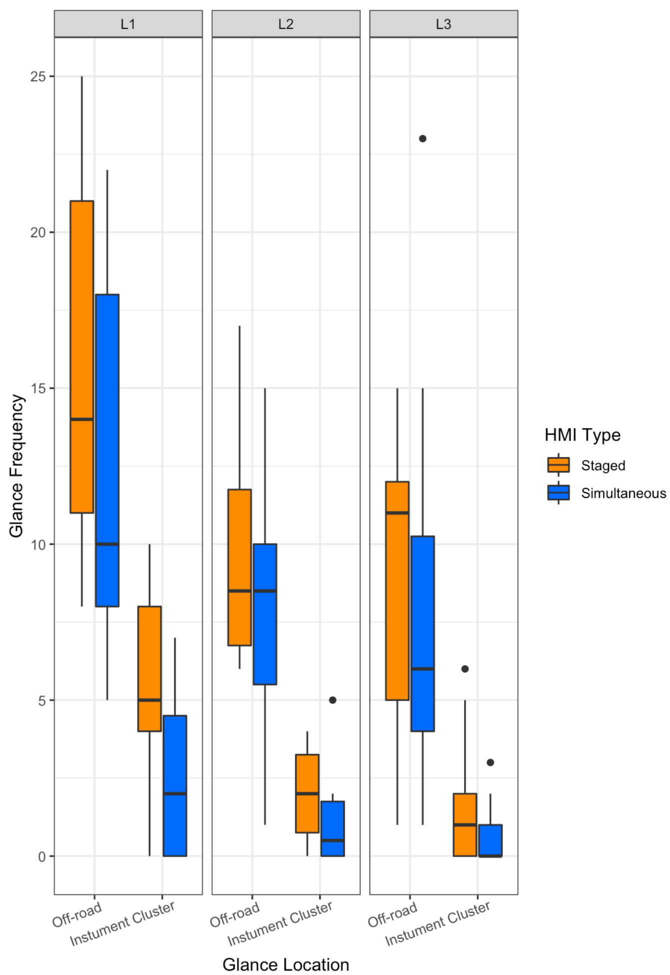


Figure 14: Boxplot of glance frequency by level of automation, glance location, and HMI type. Individual dots represent outliers.

System Usability and Trust

HMI type was the only independent variable used for the ANOVAs on the aggregate SUS score and the Trust score. No significant effect of HMI type was found for either metric. The SUS score distribution and the Trust score are shown in Figure 15 and Figure 16, respectively.

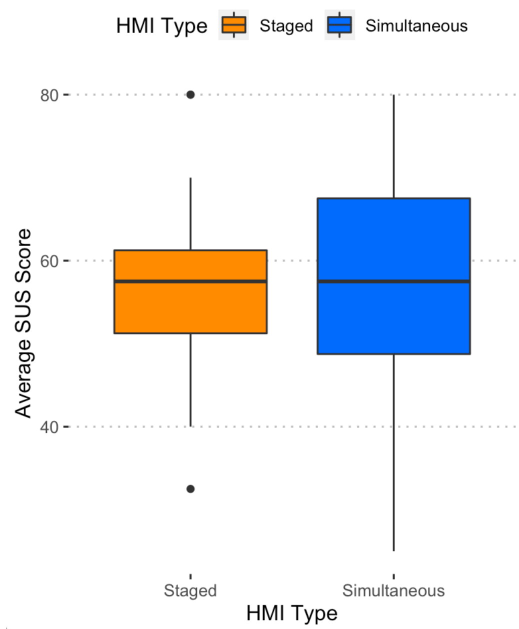


Figure 15: Average SUS score for each HMI type. The dots represent outliers.

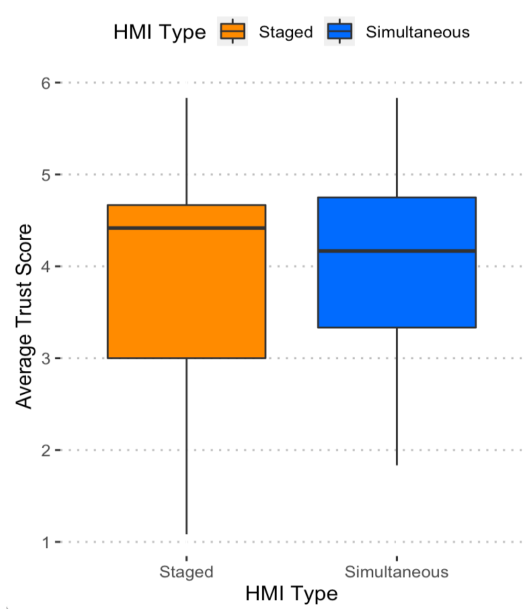


Figure 16. Average Trust score for HMI type

Discussion

The objective of this study was to examine the effectiveness of two HMIs, which vary by specificity and timing, on driver performance, eye glance behavior, and their subjective ratings of trust and usability. A simulator study was employed to assess the effectiveness of the HMIs under different levels of automation, L1 to L3. The first (Staged HMI) provided drivers with a succinct visual warning first, followed by an auditory beep five seconds later, whereas the second (Simultaneous HMI) provided drivers with a more detailed auditory and visual message simultaneously.

The first hypothesis was focused on the effect of HMI on driver performance in terms of takeover time, velocity, acceleration, proportion of large deceleration events, and standard deviation of lane position. In general, the Simultaneous HMI yielded nominally shorter response times at each level of automation compared to the Staged HMI; however, these results did not reach conventional levels of statistical significance. There was also no effect of HMI on driver velocity, but there was an interaction between HMI and time for standard deviation of velocity in the L1 drive: those who were exposed to the Staged HMI had large variations in their velocity after the takeover (in comparison to the Simultaneous HMI group). In the L1 automation drive, there was an interaction between HMI and time for acceleration, maximum negative acceleration, and proportion of large deceleration events. Taken together, it implied that those who received a Simultaneous HMI had smoother control transitions than those in the Staged HMI group, who exhibited more drastic deceleration after the takeover. The effect of HMI for acceleration could potentially indicate the existence of confounding variables (e.g., those in the simultaneous group were naturally more attentive and attuned to the environment and hence, more likely to compensate by accelerating when taking over).

The second hypothesis was that the Simultaneous HMI would result in fewer and shorter off-road glances. This was partially supported: the Simultaneous HMI resulted in fewer off-road glances and fewer glances to the instrument cluster than the Staged HMI for all levels of automation, although the effect was statistically significant only for the L3 automation drive. However, there was no effect of HMI on glance duration, either on- or off-road, or the ratio of long glances.

Finally, for the third hypothesis, there was no effect of HMI on either the usability survey or trust scores, implying that the two HMIs produced equivalent levels of trust and usability.

Overall, the current results demonstrate some benefits of the Simultaneous HMI over the Staged HMI; however, such benefits were limited in certain respects (as noted above). This could be due to a variety of reasons. First, both HMIs employed in the current study leveraged different design recommendations from past literature (e.g., Mehrotra et al., 2022). For example, the Simultaneous HMI offered more direct and clear guidance to drivers regarding appropriate decision-making, whereas the Staged HMI offered multi-modal alerts that occurred in a timed or phased manner. These varying design elements might have benefitted drivers equally or helped offset differences that would otherwise be observed. It stands to reason that a well-designed HMI that considers guidelines that are grounded in human factors principles and/or past research, should afford drivers performance benefits, while also finding them to be trustworthy and usable.

Limitations and Future Work

First, this experiment was conducted in a driving simulator. More research in the field is necessary to validate these results in real-world driving contexts. Along these lines, the current evaluation utilized only a very limited number of use case scenarios; a more comprehensive assessment in a broader array of situations is merited. Second, the participant population for this experiment skewed young and male. A more diverse participant population that is more representative of U.S. drivers is necessary. Third, though we designed the two HMIs to be distinctly different according to specificity and timing, there were only small differences between the two HMIs. A deeper analysis of other driving and behavioral metrics may yield differences between the HMIs. It is also important to note that the current results are specific to the current implementation of the HMIs and should not considered a robust assessment of the principles or guidelines that informed the implementation.

Conclusion

Past research has highlighted the importance of good HMI design for driving automation systems, but there have been limited efforts to look at how different design manifestations—modality, timing, and specificity of warnings—impact driver performance. The current study sought to examine how these elements of HMI design, implemented in different ways, affected driver behavior in terms of driving behavior, eye glances, and subjective ratings. The driving simulator experiment found a small effect of HMI on performance, suggesting benefits for multimodal presentation of alerts that support driver decision-making. Implications from this study include the need for a more thorough look into which aspects of HMI design truly affect driver behavior—especially given the array of principles that support good HMI design. In addition, the differing results across automation levels indicate the importance of considering the system or level of automation in the implementation of HMI as well as appropriately training (and warning) drivers to the abilities and limitations of driving automation systems.

References

Atwood, J. R., Guo, F., & Blanco, M. (2019). Evaluate Driver Response to Active Warning System in Level-2 Automated Vehicles. *Accident Analysis & Prevention*, *128*, 132–138. https://doi.org/10.1016/j.aap.2019.03.010

Bartlett, J. S. (2021, November 4). *How Much Automation Does Your Car Really Have?* Consumer Reports. https://www.consumerreports.org/automotive-technology/how-much-automation-does-your-car-really-have-level-2-a3543419955/

Brandenburg, S., & Roche, F. (2020). Behavioral changes to repeated takeovers in automated driving: The drivers' ability to transfer knowledge and the effects of takeover request process. *Transportation Research Part F: Traffic Psychology and Behaviour*, *73*, 15–28. https://doi.org/10.1016/j.trf.2020.06.002

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability Evaluation in Industry*. Taylor & Francis.

Cannon, B. R. (2017). *Strategies to Assist Drivers in Remaining Attentive While Under Partially Automated Driving: Verification of SuperCruise Human Machine Interface Concepts*. *2016*, 1–16. https://doi.org/10.3141/2663-03

Cohen-Lazry, G., Katzman, N., Borowsky, A., & Oron-Gilad, T. (2019). Directional tactile alerts for take-over requests in highly-automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 217–226. https://doi.org/10.1016/j.trf.2019.07.025

Cortens, B., Nonnecke, B., & Trick, L. M., (2019). *Effect of Alert Presentation Mode and Hazard Direction on Driver Takeover From an Autonomous Vehicle*, 133–139. https://doi.org/10.17077/drivingassessment.1686

Dunn, N., Dingus, T., & Soccolich, S. (2019). *Understanding the Impact of Technology: Do Advanced Driver Assistance and Semi-Automated Vehicle Systems Lead to Improper Driving Behavior?* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

Fitch, G. M., Hankey, J. M., Kleiner, B. M., & Dingus, T. A. (2011). Driver comprehension of multiple haptic seat alerts intended for use in an integrated collision avoidance system. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(4), 278–290. https://doi.org/10.1016/j.trf.2011.02.001

Friedrichs, T., Ostendorp, M. C., & Lüdtke, A. (2016). Supporting drivers in truck platooning: Development and evaluation of two novel human-machine interfaces. *AutomotiveUI 2016 - 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Proceedings*, 277–284. https://doi.org/10.1145/3003715.3005451

Gaspar, J. G., Schwarz, C., Kashef, O., Schmitt, R., & Shull, E. (2018). *Using Driver State Detection in Automated Vehicles*. https://doi.org/10.7910/DVN/FTUNZD, Harvard Dataverse, V1

He, D., Kanaan, D., & Donmez, B. (2021). In-vehicle displays to support driver anticipation of traffic conflicts in automated vehicles. *Accident Analysis & Prevention*, *149*(105842). https://doi.org/10.1016/j.aap.2020.105842

Hock, P., Kraus, J., Walch, M., Lang, N., & Baumann, M. (2016). Elaborating feedback strategies for maintaining automation in highly automated driving. *AutomotiveUI 2016 - 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Proceedings*, 105–112. https://doi.org/10.1145/3003715.3005414

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Kasuga, N., Tanaka, A., Miyaoka, K., & Ishikawa, T. (2020). Design of an HMI System Promoting Smooth and Safe Transition to Manual from Level 3 Automated Driving. *International Journal of Intelligent Transportation Systems Research*, *18*(1), 1–12. https://doi.org/10.1007/s13177-018-0166-6

Kim, J., Kim, W., Kim, H. S., & Yoon, D. (2018). Effectiveness of Subjective Measurement of Drivers' Status in Automated Driving. *IEEE Vehicular Technology Conference*, *2018-Augus*, 18–19. https://doi.org/10.1109/VTCFall.2018.8690557

Lee, J., & Yang, J. H. (2020). Analysis of driver's EEG given take-over alarm in SAE level 3 automated driving in a simulated environment. *International Journal of Automotive Technology*, *21*(3), 719–728. https://doi.org/10.1007/s12239-020-0070-3

Li, S., Blythe, P., Guo, W., Namdeo, A., Edwards, S., Goodman, P., & Hill, G. (2019). Evaluation of the effects of age-friendly human-machine interfaces on the driver's takeover performance in highly automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *67*, 78–100. https://doi.org/10.1016/j.trf.2019.10.009

Louw, T., Merat, N., Jamson, H., & University of Iowa, I. C. (2015). Engaging with Highly Automated Driving: To be or Not to be in the Loop?, *Driving Assessment Conference 8*(2015), 190–196. https://doi.org/10.17077/drivingassessment.1570

Mehrotra, S., Wang, M., Wong, N., Parker, J., Roberts, S. C., Kim, W., Romo, A., & Horrey, W. J. (2022). *Human-Machine Interfaces and Vehicle Automation: A Review of the Literature and Recommendations for System Design, Feedback, and Alerts* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

Mok, B., Johns, M., Yang, S., & Ju, W. (2017). Actions speak louder: Effects of a transforming steering wheel on post-transition driver performance. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 1–8. https://doi.org/10.1109/ITSC.2017.8317878

Nakashima, A., & Crébolder, J. M. (2010). Evaluation of audio and visual alerts during a divided attention task in noise. *Canadian Acoustics*, *38*(4), 3–8. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2317

Petermeijer, S., Doubek, F., & de Winter, J. (2017, October 5). Driver Response Times to Auditory, Visual and Tactile Take-Over Requests: A Simulator Study with 101 Participants. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1505–1510. https://doi.org/10.1109/SMC.2017.8122827.

Roche, F., & Brandenburg, S. (2018). Should the urgency of auditory-tactile takeover requests match the criticality of takeover situations. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 1035–1040. https://doi.org/10.1109/ITSC.2018.8569650

SAE. (2021). Taxonomy and Definitions for terms related to driving automation systems (Standard No. J3016). In *SAE International*. Society of Automotive Engineers. https://www.sae.org/standards/content/j3016_202104/

Telpaz, A., Rhindress, B., Zelman, I., & Tsimhoni, O. (2015). Haptic Seat for Automated Driving: Preparing the Driver to Take Control Effectively. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*, 23–30. https://doi.org/10.1145/2799250.2799267

van der Heiden, R. M. A., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. *Conference on Human Factors in Computing Systems - Proceedings*, *2017-May*, 392–404. https://doi.org/10.1145/3025453.3025507

Vogelpohl, T., Kühn, M., Hummel, T., Gehlert, T., & Vollrath, M. (2018). Transitioning to Manual Driving Requires Additional Time after Automation Deactivation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *55*, 464–482. https://doi.org/10.1016/j.trf.2018.03.019

Yang, Y., Karakaya, B., Dominioni, G. C., Kawabe, K., & Bengler, K. (2018). An HMI Concept to Improve Driver's Visual Behavior and Situation Awareness in Automated Vehicle. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2018-Novem*, 650–655. https://doi.org/10.1109/ITSC.2018.8569986

Yoon, S. H., Kim, Y. W., & Ji, Y. G. (2019). The effects of takeover request modalities on highly automated car control transitions. *Accident Analysis & Prevention*, *123*, 150–158. https://doi.org/10.1016/j.aap.2018.11.018

Yun, H., & Yang, J. H. (2020). Multimodal warning design for take-over request in conditionally automated driving. *European Transport Research Review*, *12*(1). https://doi.org/10.1186/s12544-020-00427-5

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis and Prevention*, *78*, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

Appendix A

The p-values for all measures of driving and takeover performance/behavior are summarized in Table 6 and eye glance behavior in Table 7.

Table 6. p-values of the statistical analysis for takeover time, velocity, standard deviation of velocity, acceleration, minimum acceleration, time proportion of large deceleration, and standard deviation of lane position; cell values with an asterisk represents significance.

	Level of Automation	HMI Type	Time	HMI Type * Time
Takeover Time	L1	0.06	-	-
	L2	0.19	-	-
	L3	0.36	-	-
Velocity	L1	0.83	0.12	0.36
	L2	0.39	0.12	0.69
	L3	0.43	< .001*	0.97
SD of Velocity	L1	0.31	< .001*	0.02*
	L2	0.22	< .001*	0.53
	L3	0.88	< .001*	0.98
Acceleration	L1	0.99	< .001*	< .001*
	L2	0.54	< .001*	0.51
	L3	0.98	0.02*	0.82
Max. (Neg.) Acceleration	L1	0.29	< 0.01*	< .001*
	L2	0.72	0.35	0.57
	L3	0.31	< .001*	0.79
Time Proportion of Large Deceleration	L1	0.68	0.71	0.02*
	L2	0.64	0.96	0.97
	L3	0.44	0.02*	0.41
SD of Lane Position	L1	0.39	< 0.01*	0.80
	L2	0.61	< .001*	0.72
	L3	0.80	< .001*	0.36

Table 7. p-values of the glance data analysis with HMI type being the independent variable; cell values with an asterisk represents significance.

	L1	L2	L3
Mean Glance Duration			
On-road	0.28	0.96	0.35
Off-road	0.53	0.76	0.38
Towards instrument cluster	0.55	0.56	0.99
Total Glance Duration			
Off-road	0.28	0.76	0.47
Towards instrument cluster	0.25	0.98	0.12
Glance Frequency			
Off-road	0.32	0.40	< .001*
Towards instrument cluster	0.13	0.48	0.02*
Long-Glance Ratio			
Off-road	1.00	1.00	0.86