



The effects of takeover request lead time on drivers' situation awareness for manually exiting from freeways: A web-based study on level 3 automated vehicles

Xiaomei Tan, Yiqi Zhang^{*}

Department of Industrial and Manufacturing Engineering, Pennsylvania State University-University Park, State College, PA, United States

ARTICLE INFO

Keywords:

Freeway exiting
Takeover request
Lead time
Situation awareness
Conditionally automated driving

ABSTRACT

Conditional automation systems allow drivers to turn their attention away from the driving task in certain scenarios but still require drivers to gain situation awareness (SA) upon a takeover request (ToR) and resume manual control when the system is unable to handle the upcoming situation. Unlike time-critical takeover situations in which drivers must respond within a relatively short time frame, the ToRs for non-critical events such as exiting from a freeway can be scheduled way ahead of time. It is unknown how the ToR lead time affects driver SA for resuming manual control and when to send the ToR is most appropriate in non-critical takeover events. The present study conducted a web-based, supervised experiment with 31 participants using conditional automation systems in freeway existing scenarios while playing a mobile game. Each participant experienced 12 trials with different ToR lead times (6, 8, 10, 12, 14, 16, 18, 20, 25, 30, 45, and 60 s) for exiting from freeways in a randomized order. Driver SA was measured by using a freeze probe technique in each trial when the participant pressed the spacebar on the laptop to simulate the takeover action. Results revealed a positive effect of longer ToR lead times on driver SA for resuming control to exit from freeways and the effect leveled off at the lead time of 16–30 s. The participants tended to postpone their takeover actions further when they were given a longer ToR lead time and it did not level off up to 60 s. Nevertheless, not all drivers waited till the last moment to take over AVs even though they did not get sufficient SA. The ToR lead time of 16–30 s was recommended for better SA; and it could be narrowed down to 25–30 s if considering the subjective evaluations on takeover readiness, workload, and trust. The findings provide implications for the future design of conditional automation systems used for freeway driving.

1. Introduction

The conditional driving automation or Level 3 automated vehicles (AVs; SAE International, 2016) enable drivers to turn their attention away from the driving task in certain traffic situations. However, drivers are still required to serve as a fallback for automation and to intervene upon the request of the car when it is unable to handle the forthcoming situation. One typical driving situation that limits the current Level 3 AVs is navigating freeway/highway/motorway off-ramps, which are usually one-way, steeply curved, and banked road segments (Thorn et al., 2018). While the automated driving on freeways appears to be the first ready-to-market feature of Level 3 AVs, a control transition from automation systems to the human driver is still required for exiting a freeway (Holländer and Pfleging, 2018).

In the process of control transitions, the lead time of takeover request (ToR) has a direct effect on how much situation awareness (SA) can be obtained by drivers (Lu et al., 2017; Pampel et al., 2019; Samuel et al., 2016). The ToR lead time is defined as the time budget for drivers retrieving control of the car before arriving at the desired exit, namely the time duration from the presentation of ToR to the car arriving at the beginning of the exit if the driver did not intervene and let the car maintain the current automated driving speed (Nobari et al., 2020; Wörle et al., 2020). The following sections provide a review of existing empirical studies on the effects of ToR lead time on driver SA in non-scheduled takeovers and the effects of ToR characteristics on takeover performance in freeway exiting takeovers as well as research objectives and hypotheses of the present study.

^{*} Corresponding author at: Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, Pennsylvania State University, University Park, State College, PA 16801, United States.

E-mail address: yuz450@psu.edu (Y. Zhang).

<https://doi.org/10.1016/j.aap.2022.106593>

Received 28 June 2021; Received in revised form 21 December 2021; Accepted 30 January 2022

Available online 15 February 2022

0001-4575/© 2022 Elsevier Ltd. All rights reserved.

1.1. Effects of ToR lead time on driver SA in non-scheduled takeovers

Situation awareness was defined by Endsley (1988) as a person's perception of the elements of the environment (Level 1 SA), the comprehension of their meaning (Level 2 SA), and the projection of their future status (Level 3 SA). It is an important human factors construct worth studying in the context of conditional driving automation, as the driver who has little to no SA in automated mode needs to acquire SA and decision making for the purpose of executing takeover maneuvers safely (Forster et al., 2017; Miller et al., 2014). Drivers' ability to retrieve control of Level 3 AVs is positively affected by an increase in driver SA (Van den Beukel and Van der Voort, 2013b).

For unexpected, non-scheduled takeover situations (e.g., emergency brake in congestion and system limit), existing studies have found that drivers had better SA when they were given longer ToR lead times, which brought a higher chance for successful takeovers or enhanced drivers' ability to anticipate latent hazards on the road (Samuel et al., 2016; Van den Beukel and Van der Voort, 2013a, 2013b; Vlakveld et al., 2018; Wright et al., 2016). Specifically, Van den Beukel and Van der Voort (2013a, 2013b) concluded that the ToR with 2.8 s lead time yielded better self-reported SA and a higher success rate of takeover than that in 1.5 s and 2.2 s ToR conditions for handling urgent takeover events on congested freeways. Samuel et al. (2016) and Wright et al. (2016) tested driver SA under 4–12 s ToR conditions and found that 8 s lead time was sufficient for drivers to build up SA and detect static latent hazards that were statistically equivalent to that in manual driving. However, when some latent hazards were dynamic, there was a decline in the proportion of hazards detected (Vlakveld et al., 2018), which suggested that a ToR lead time longer than 8 s might be required for drivers to achieve sufficient SA for detecting dynamic latent hazards. These diverse findings suggest that a relation between the ToR lead time and driver SA for a specific takeover situation would exist; the relations could be different for various takeover situations as the allowed time frames for drivers to respond may vary.

1.2. Effects of ToR characteristics on takeover performance in freeway exiting takeovers

Previous studies that investigated scheduled, non-time-critical takeovers (e.g., freeway exiting and entering) were mostly focused on drivers' takeover

performance and subjective experience. Table 1 presented a summary of literature that involved freeway exiting takeovers. Some studies have quantified the effects of ToR characteristics such as ToR modality and visual representation on the takeover performance (Holländer and Pfleging, 2018; Langlois and Soualmi, 2016; Petermeijer et al., 2017; Yun and Yang, 2020). However, there are few studies focusing on the design of ToR lead time for freeway exiting takeovers.

The ToR for exiting a freeway can be scheduled to notify the driver at a relatively early point (McCall et al., 2019). The existing studies on freeway exiting takeovers have been focused on investigating drivers' takeover performance with the adopted ToR lead time ranging from 10 s to 60 s, but few studies offered insight into the effects of ToR lead time on driver SA for freeway exiting takeovers. One exception was a simulator study by Pampel et al. (2019) that found the longer ToR lead time helped drivers build up more appropriate SA and enabled better longitudinal vehicle control after takeover. However, Pampel et al. (2019) only compared two ToR lead times, namely planned, 50 s ToRs for freeway exiting and unplanned, 5 s ToRs caused by system limit. It is unclear how driver SA and takeover performance would change with the ToR lead time between 5 s and 50 s. Another study that investigated driver SA used video clips of different lengths (1, 3, 7, 9, 12, and 20 s) to denote the time duration between the presentation of a ToR and the takeover action, but it did not include freeway exiting takeovers and it did not require any takeover operations (Lu et al., 2017). Participants were asked to view the video clips of freeway driving and to reproduce the traffic situation at the end of each video. Results showed positive effects of the longer video length on driver SA in terms of the reproduction of situation layout, and the effects saturated at 7–12 s; the assessment of relative speeds improved with the video length, and it did not show saturation up to 20 s. The results suggest the necessity of exploring ToR lead times longer than 20 s in freeway exiting scenarios when it involves the takeover process and allows drivers' self-paced takeover decisions. To the best of the authors' knowledge, no existing studies have examined the appropriateness of ToR over a large range of lead time for freeway exiting takeovers in Level 3 AVs from multiple perspectives such as driver SA, takeover performance, and subjective evaluations on ToR systems.

1.3. Research objectives and hypotheses

The objective of the present study is to investigate the association

Table 1
Summary of empirical studies involving freeway exiting takeover scenarios.

No.	Literature	Takeover situation	ToR modality	ToR lead time	Takeover reaction time	Driver SA
1	Pampel et al. (2019)	Exit highways, missing lane markings, traffic dispersal	V	50 s (exit highways); 5 s (system limit events)	Not reported	Better SA following 50 s ToRs compared with 5 s ToRs
2	Holländer & Pfleging (2018)	Exit highways	A + V	Visual pre-warning (60 s) + ToR (<10 s)	0.53–8.89 s ($M = 2.78$)	Not reported
3	Langlois & Soualmi (2016)	Change lanes to exit highways and deal with a braking vehicle ahead	A + V	30 s	$M = 5.09$ s ($SD = 2.58$)	Not reported
4	Metz et al. (2020)	Exit highways, construction sites, adverse weather conditions, missing lane markings	–	15 s	3–4 s	Not reported
5	Nobari et al. (2020)	Change lanes to exit highways	A + V	18 s	$M = 4.21$ s ($SD = 1.86$)	Not reported
6	Petermeijer et al. (2017)	Exit motorways, breakdown vehicles, lane closed, roadworks, traffic jam	A, V, T	Greater than 10 s	V: $M = 2.29$ s ($SD = 1.71$) A: $M = 1.54$ s ($SD = 0.63$) T: $M = 1.47$ s ($SD = 0.49$)	Not reported
7	Wörle et al. (2020)	Change lanes to exit motorways, roadworks	A + V	Multi-stage ToRs at 60 s, 15 s, 11 s, 7 s, and emergency stop	3.3–50.4 s ($M = 18.8$, $SD = 16.9$)	Not reported
8	Yun & Yang (2020)	Change lanes to exit motorways	A + V, A + V + T	15 s	A + V: $M = 2.28$ s ($SD = 0.71$) A + V + T: $M = 2.01$ s ($SD = 0.55$)	Not reported

Note: A = auditory; V = visual; T = tactile.

between ToR lead time and driver SA at the moment when drivers disengage the automation mode and take over control to exit from freeways in Level 3 AVs. Drivers become less situationally aware when they are out of loop in Level 3 AVs than they are when drive in full control (Endsley, 1995; Gugerty, 1997; Samuel et al., 2016). To regain SA and prepare for the upcoming control transition, drivers may prefer a longer ToR lead time so that they can have sufficient time to assess the situation before takeover (Lu et al., 2017). The scheduled takeover for exiting a freeway makes the long ToR lead time feasible. Nevertheless, it is unknown whether the positive effects of longer ToR lead times on driver SA and subjective evaluations would saturate with the increase of ToR lead time for freeway exiting. It is also unknown how drivers' takeover reaction time would change with the ToR lead time. When given different amounts of time to resume control and exit from freeways, drivers may adopt different strategies to execute the takeover decision based on the situation criticality and their SA. It would raise the risk of traffic accidents when a driver takes control of the car without yet regaining appropriate SA. Knowing how drivers would make self-paced takeover decisions based on their SA under different ToR lead time conditions will contribute to assessing drivers' readiness for takeover at a certain time and determining whether it is safe for the automation systems to hand over control to the driver. Therefore, the present study aims to fill the research gap by investigating the effects of ToR lead time on driver SA when executing the takeover action, takeover reaction time, and subjective evaluations and exploring the optimal ToR lead time for freeway exiting takeovers (Fig. 1).

The hypothesized trends of driver SA and takeover reaction time with the increase of ToR lead time for freeway exiting are illustrated as Fig. 1. The endpoint t_0 is the minimum time for drivers acquiring motor readiness and redirecting gaze. The other endpoint t_{SA} is the minimum time that drivers need to gain sufficient SA for resuming control and exiting a freeway. The difference between t_0 and t_{SA} (i.e., $t_{SA} - t_0$) represents the required time taken for scanning, obtaining cognitive readiness, and selecting actions for successful takeovers. Three areas denoted by different colors are generated by the two vertical dash lines at t_0 and at t_{SA} : The area ① represents freeway exiting failures caused by the insufficient lead time for gaze redirection and motor readiness when the ToR lead time is shorter than t_0 . The area ② represents that drivers are able to deactivate the automated mode successfully with insufficient cognitive readiness for takeover and for the action selection after

takeover, when the ToR lead time is between t_0 and t_{SA} . In other words, driver SA at the moment of takeover might not be good enough for ensuring stable driving performance after takeover, which raises the risk of road accidents. In area ②, driver SA is hypothesized to increase steeply with the ToR lead time. In this stage, drivers may prefer to execute takeover action at the very last moment before exiting to obtain SA as much as they can. Therefore, the takeover reaction time is hypothesized to approximate the given ToR lead time with the slope of takeover reaction time being about 1. In area ③, when a ToR with lead time longer than t_{SA} is provided, it is hypothesized that the slope of driver SA starts decreasing and driver SA levels off at a certain point. As a result of sufficient SA obtained, drivers may postpone their takeover action intentionally, but not necessarily to wait till the last moment before arriving at the exit. Thus, the slope of increasing takeover reaction time with ToR lead time is hypothesized to be smaller than 1 in area ③. As the time budget given to the driver increases to some point, we hypothesize that there is a higher likelihood that drivers choose to play the game and wait till a late time point to take over. The distraction may lead to a downtrend of driver SA in comparison to the shorter ToR lead time conditions when drivers choose to monitor the traffic constantly (Heikoop et al., 2018; Yang et al., 2020). The present study is focused on areas ② and ③ and examines the hypotheses on the trends of driver SA and takeover reaction time with the increase of ToR lead time for freeway exiting takeovers.

2. Method

2.1. Participants

A total of 31 subjects participated in the experiment. Thirty participants (12 males, 17 females, and 1 unspecified) remained for the analysis after removing 1 participant's invalid data because of technical issues. Participants were aged between 21 and 53 years old ($Mean = 28.8$, $SD = 8.4$). All were licensed drivers with an average driving experience of 11.2 years ($SD = 8.3$) and an average mileage of 7,860 miles (12, 650 km) per year ($SD = 5,158.4$ miles (8,301.6 km)). A summary of descriptive statistics was shown in Table 2. The participants were recruited from the general public in the United States via Penn State's StudyFinder website and a participant recruitment group on Facebook. Each participant received a \$20 eGift Card as compensation.

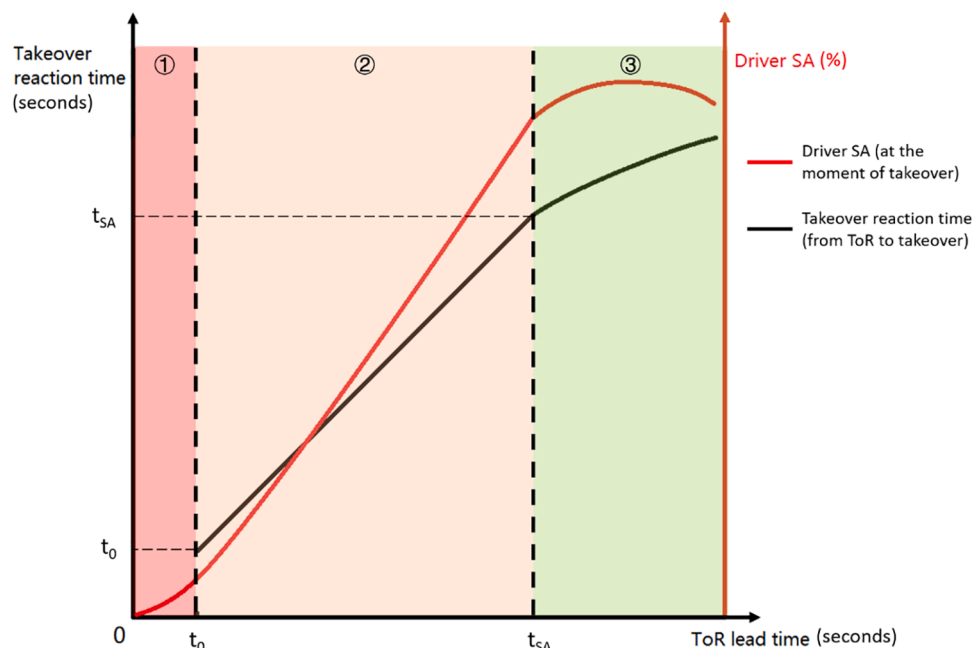


Fig. 1. Hypothesized trends of driver SA and takeover reaction time with ToR lead time.

Table 2
Descriptive statistics of demographics.

Variables	Descriptive Statistics (N = 30)
Age	21 ~ 53 (<i>Mean</i> = 28.8, <i>SD</i> = 8.4)
Gender	Males (40%); Females (56.7%); Unspecified (3.3%)
Racial	White and non-Hispanic (76.7%) Black and non-Hispanic (10%) Asian (10%) Unspecified (3.3%)
Annual mileage	500 ~ 17,000 (<i>Mean</i> = 7,860, <i>SD</i> = 5,158.4)
Years of driving	3 ~ 38 (<i>Mean</i> = 11.2, <i>SD</i> = 8.3)
Freeway driving frequency (per year)	More than 50 times (46.7%) 41 ~ 50 times (0%) 31 ~ 40 times (10%) 21 ~ 30 times (16.7%) 11 ~ 20 times (20%) 1 ~ 10 times (6.7%) Never (0%)

2.2. Experiment design

The takeover process of exiting a freeway in Level 3 AVs is illustrated as Fig. 2. After a ToR is initiated by the automation system, drivers first redirect their gaze from non-driving activities to the driving task. Then, they scan the roadway and car dashboard to build up SA (e.g., takeover task, roadway, car positions, surrounding vehicles, driving speed, speed limit). In the meanwhile, drivers establish the cognitive and motor readiness for takeover. Finally, drivers execute the takeover decision, retrieve control of the car, and exit from the freeway manually.

The experiment adopted a within-subjects design with the ToR lead time as the only independent variable. The ToR lead time was defined as the distance away from the desired exit when the ToR was delivered divided by the automated driving speed of the subject vehicle. As the ToR for freeway exiting could be scheduled ahead of time, a large range of ToR lead time including 12 levels (i.e., 6, 8, 10, 12, 14, 16, 18, 20, 25, 30, 45, and 60 s) was tested in a randomized order. The shortest lead time (6 s) represented the minimum time required by the audio ToR message to be completely delivered. The 60 s was selected as the highest level as it was the longest ToR lead time that has been used for freeway exiting takeovers in literature (see Table 1).

The driving scenarios were created in the driving simulator STISIM Drive® M300WS-Console system. The driving context was freeways, which were designed exclusively for high-speed vehicular traffic. No traffic lights, intersections, railways, or pedestrian paths were included.

The traffic was light with an average traffic density around 12 vehicles per mile per lane (7 vehicles/km/lane). The subject vehicle was in automation mode at the beginning of each scenario and all of surrounding vehicles drove at the speed in accordance with speed limits varying from 50 to 60 mph (80–97 km/h). Each freeway scenario was 17,000–18,000 feet long (5.2–5.5 km; about 3.5 min' driving). The freeway exit signs were placed at 2 miles, 1 mile, and right at the exit gore area. A ToR was sent out when the *remaining time to exit* (see Fig. 2), which was defined as the time left until the subject vehicle arrived at the beginning of the desired exit based on its speed and the distance away from the exit, was equal to the designated ToR lead time.

To reduce the carryover effect caused by the within-subjects design, we created 12 driving scenarios across 12 trails for each participant by changing the types of freeway exits, the number of lanes, and the side of the freeway an exit was on with the purpose to increase the variance of scenarios. Examples of scenarios are shown in Table 3. Two types of freeway exits were designed in the experiment – exit only and not exit only. It was assumed that a car driving in an exit-only lane needs to swerve when the exit lane is separated from the original freeway, whereas a car driving in a not exit-only lane needs to swerve at the beginning of the exit ramp if it intends to exit from the freeway. There were either 2 or 3 or 4 lanes in the subject vehicle's driving direction. Based on the assumption that the automated vehicles driving on a freeway would be able to change lanes automatically with a navigation route being entered, the subject vehicle in the experiment had changed to the exit lane before the ToR was issued in each trial so as to get rid of the bias introduced by the extra time required for manual lane changes. In each scenario, there was a car driving 300 feet (91 m) ahead of the subject vehicle in the same lane and a car driving 250 feet (76 m) ahead of the subject vehicle in the adjacent lane, which are denoted by Car 1 and Car 2 in Table 3, respectively. At most one of the two cars in a scenario had activated the turn signals to exit (for Car 1) or to change lanes to exit (for Car 2) when the ToR was issued. The freeway exit maneuvers of Car 1 and Car 2 were always performed at the last moment before arriving at the exit, which ensured the SA queries (see section 2.5) were applicable for all situations regardless of when participants chose to take over.

The scenarios of freeway exit on the left or the right side of the road were evenly balanced. In half of trials (6, 10, 14, 18, 25, 45 s), the exit lane was exit only in which cars typically did not need to signal their intention to exit. The other half of trials (8, 12, 16, 20, 30, 60 s) included a 300 feet (91 m) long exit ramp that branched off from the original lane. When the subject vehicle could either exit or remain driving on the

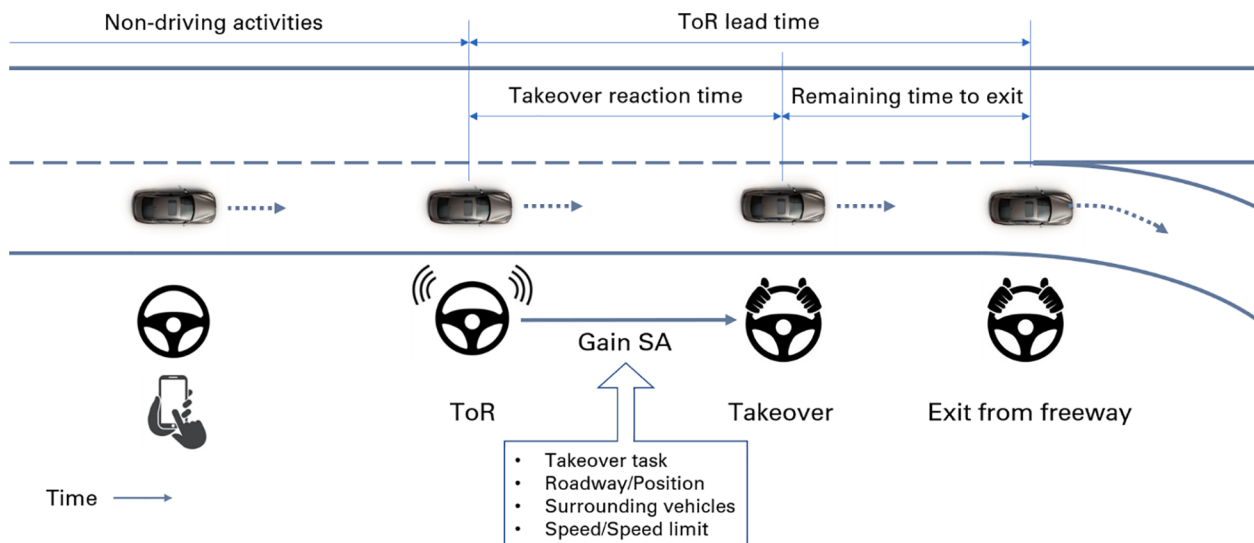
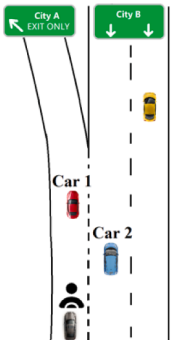
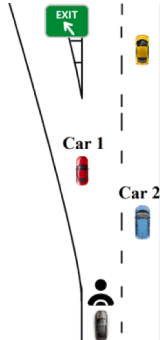
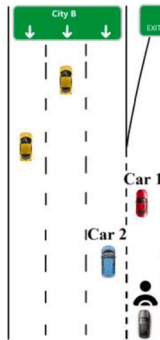
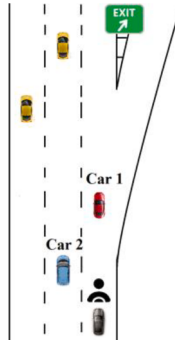
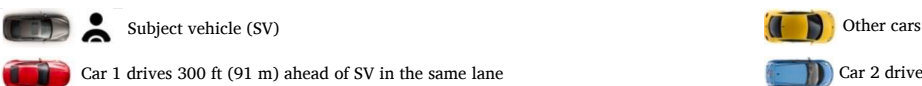


Fig. 2. The takeover process of exiting a freeway in Level 3 AVs.

Table 3
Examples of driving scenarios in video stimuli.

ToR lead time	6 s	16 s	45 s	60 s
Diagram				
Exit type	Exit only	Not exit only	Exit only	Not exit only
Turn signal	Car 1, Car 2, and SV have their turn signal off.	Car 1 has had its left turn signal on since the ToR. Car 2 has its turn signal off. SV activates its left turn signal 10 s prior to arriving at the exit ramp.	Car 2 has its right turn signal on since ToR. Car 1 and SV have their turn signal off.	Car 1 has its turn signal off. Car 2 has its right turn signal on since ToR. SV activates its right turn signal 10 s prior to arriving at the exit ramp.
Speed limit	SV drives at 50 mph (80 km/h). Speed limit to exit was 40 mph (64 km/h).	SV drives at 50 mph (80 km/h). Speed limit to exit was 50 mph (80 km/h).	SV drives at 60 mph (97 km/h). Speed limit to exit was 45 mph (72 km/h).	SV drives at 60 mph (97 km/h). Speed limit to exit was 50 mph (80 km/h).
ToR message	"In 0.1 mile, exit from freeway on the left side to DeWitt. Please take over!"	"In 0.26 mile, exit from freeway on the left side to San Diego. Please take over!"	"In 0.75 mile, exit from freeway to Morristown Mahwah. Please take over!"	"In 1 mile, exit from freeway to Danbury. Please take over!"
				
Car 1 drives 300 ft (91 m) ahead of SV in the same lane			Car 2 drives 250 ft (76 m) ahead of SV in the adjacent lane	

freeway, it was required to activate the turn signals if it intended to exit. In this experiment, the subject vehicle in automation mode would turn on the signal lights automatically 10 s prior to exiting a freeway. With that being said, the turn signal was already on when the ToR was issued 10 s or shorter before arriving at the exit. Whereas in the trials with ToR lead times longer than 10 s, the subject vehicle remained its turn signal off when the ToR was issued; then it either activated its turn signal if the participant did not press the spacebar till the last 10 s or had them off otherwise. Additionally, when the ToR was given, the subject vehicle drove at a speed of either 50 or 60 mph (80 or 97 km/h), while the speed limit posted on the exit was either smaller than or the same as the subjective vehicle's speed.

2.3. Stimuli

Given the current in-person study challenges and health concerns due to COVID-19, a web-based experiment was developed and conducted using Gorilla Experiment Builder (gorilla.sc; Anwyl-Irvine et al., 2020). The material used in the experiment was produced by recording the main display screen of the driving simulator while it was running each scenario file. To reduce the strain on participants' internet connection to download and the delay in data capturing, rather than using the complete recorded video of each scenario, we created a video clip by trimming the start of the video right before the onset of ToR. Before such a video stimulus was given, a series of screenshots captured from the recorded video were presented considering the automated driving prior to ToRs did not require participants' monitoring or response. A car engine sound effect (~55 dB) was played continuously to simulate the environment of riding in an automated vehicle.

An example of images (1920 × 1080 pixels) and a screenshot of the following video stimulus (1280 × 720 pixels) are shown as Fig. 3. The images that were presented prior to the ToR conveyed information such as a mockup of navigation map, speedometer, and odometer as shown in the left column of Fig. 3. Along with the image, the car engine sound effect was played with a duration of 6 s or 15 s. The image automatically advanced and switched to the next one once the audio was finished. The

ToR comprised of an audio warning and a visual stimulus as shown in the right column of Fig. 3. The audio warning consisted of a 1 kHz warning tone lasted for 0.15 s and a following speech message. The speech was presented in a digitized human female voice with a speech rate of ~150 words/min at the loudness level of 70 dB. At the same time when the audio ToR was issued, the visual stimulus of the navigation map started displaying the information about the freeway exit, including the distance away from the exit when the ToR was issued, the speed limit at the exit ramp, and whether the lane that the subject vehicle was driving in was exit only or not. The videos stopped playing as soon as the participant pressed the spacebar to simulate deactivating the automation mode for resuming control. If the participant did not intervene, the video would reach the end when the subject vehicle passed the freeway exit.

2.4. Dependent measures

Three general categories of dependent measures are driver SA, takeover performance, and subjective evaluations on ToR systems. Table 4 presents a summary of dependent variables including units, measurement, expected trends with ToR lead time, and references.

• Driver SA

Driver SA was measured using Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1987), which is a probe recall technique. When the participant pressed the spacebar on keyboard to indicate a takeover action, the video froze and it jumped to the next web page with SA queries about the current situation (see Appendix 1). In each trial, there were 4 sets of SA queries focusing on Car 1, Car 2, SV's speed, and SV's turn signal state. Each set included three single-choice questions that corresponded to three levels of SA, namely perception (Level 1), comprehension (Level 2), and projection (Level 3).

The score for each SA query was determined by checking the participant's answer against the video recorded during the experiment, resulting in either 1 for correct answer or 0 for incorrect answer. Based on the scores, the overall response accuracy (i.e., the percentage of



Fig. 3. An example of image presented before the ToR (left) and a screenshot of the video stimulus presented after the ToR (right).

correct answers) based on all of the 12 SA queries, as well as the response accuracy for the 4 queries in each one of three SA levels were calculated for each participant.

The three-level SA queries were followed by a slider scale question to evaluate participants' distance perception in freeway existing scenarios as an additional measure of driver SA, which is an important element at the tactical level that enables the driver to form expectations on how the scenario will unfold over time and select an appropriate exit maneuver (Sukthankar, 1997). The slider scale was designed with the left end representing the location where the ToR was issued and the right end representing the freeway exit (see Question 5 in Appendix 1). The unit of the scale was 0.01 mile. Participants were asked to drag the slider to the location where they believed the subject vehicle was when the ToR was sent out. The distance between the slider and the right end of the scale was defined as the *perceived distance* (d_p) away from the exit. The *actual distance* (d_a) was calculated by subtracting the distance that the car had traveled – which was read from the odometer right before the video stopped – from the designed longitudinal distance of the exit. Based on these two measures, the *absolute error of distance* to exit was calculated according to Eq. (1). This metric was used to assess participants' awareness of the travel distance to the desired exit.

$$\text{Absolute Error of Distance} = |d_p - d_a| \quad (1)$$

• Subjective evaluations

The rest of the queries (see Questions 6 to 10 in Appendix 1) were designed to obtain the subjective evaluations on the ToR system after each trial. The measures with corresponding questions were

- readiness for takeover (“How much were you ready to take over control of the car to exit from freeway?”)

- appropriateness of ToR timing (“What do you think about the timing of the audio takeover request?”)
- workload (“How much do you trust the takeover request system?”)
- trust (“What was the workload for you in responding to the takeover request?”)
- acceptance (“How much do you accept the takeover request system?”)

Participants were asked to rate on single-item 9-point Likert scales with 1 indicating the low end, 9 indicating the high end, and 5 indicating the neutral midpoint. For example, Question 6 asked participants to evaluate their readiness for takeover using a scale ranging from 1 (not ready at all) to 5 (neutral) to 9 (fully ready).

• Takeover performance

In addition to SAGAT queries, two takeover performance measures – takeover reaction time and exiting outcome – were obtained from the experiment. The takeover reaction time starting from the ToR to the moment when the participant pressed the spacebar was automatically collected by Gorilla.sc. The exiting outcome was quantified as dichotomous data by comparing the takeover reaction time and the ToR lead time. If the participant took longer to press the spacebar than the ToR lead time, namely took over after passing the swerve-requested location, it was counted as a failure to exit from the freeway. Otherwise, it was counted as a successful exit.

2.5. Data analysis

As a within-subjects design was adopted in this study in which 12

Table 4

A summary of dependent variables.

Dependent measure		Unit	Measurement	Expected trend with ToR lead time	Reference
Driver SA	Overall/Level 1/Level 2/Level 3 SA response accuracy	%	Freeze probe technique (SAGAT) when participants executed takeover in each task	Increase with ToR lead time and then stabilize	Lu et al. (2017)
	Absolute error of distance	feet	A slider scale question during the freeze probe in each task	Increase with ToR lead time and then stabilize	
Subjective evaluations	Readiness for takeover	1	A single-item 9-point Likert scale for each measure during the freeze probe in each task	Increase with ToR lead time and then stabilize	Du et al. (2020)
	Appropriateness of ToR timing			Change from late to appropriate to early	Mok et al. (2015)
	Post-trial trust			Increase with ToR lead time and then stabilize	
	Workload			Decrease with ToR lead time and then stabilize	
	Acceptance			Increase with ToR lead time and then stabilize	Yun, Oh, & Myung (2019)
Takeover performance	Takeover reaction time	second	Automatically recorded by Gorilla.sc in each task	Increase with ToR lead time	Jin et al. (2021)
	Exiting outcome	binary bit	Determined by comparing takeover reaction time and ToR lead time in each task	Fewer exiting failures when ToR lead time was long	

trials were nested within 30 participants, the multilevel modeling test was conducted to quantify the within-person variability (i.e., how much a dependent variable differs from trial to trial within a participant) and between-person variability (i.e., how much a participant differs from others). The multilevel modeling analysis followed a general “build-up” strategy proposed by Heck et al. (2013). To start with, we examined whether there was significant variance across participants in each dependent variable for driver SA, takeover performance, and subjective evaluations. A random intercept model (i.e., an empty, unconditional model with no predictors included) was tested for each dependent measure, which was represented by the following equations:

Level 1 equation (modeling within-person variation):

$$DV_{ij} = \beta_{0j} + e_{ij} \quad (2)$$

Level 2 equation (modeling variation in intercepts):

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \quad (3)$$

where the “ij” in subscripts represents the *i* th trial for participant *j*; DV_{ij} denotes a dependent measure; β_{0j} denotes the intercept for participant *j*; e_{ij} denotes the random within-person residual (error term); γ_{00} denotes the fixed effect of the average intercept across participants; μ_{0j} denotes the variability in person-level intercepts, which is considered the random effect.

The Wald Z test was conducted and the intraclass correlation coefficient (ICC) was calculated to identify if there was significant variance across participants. If a substantial clustering was detected in data, the multilevel model was constructed with participants as higher-level units. Otherwise, a single-level general linear model was built without considering the between-person variability.

In the second step, a model was tested after incorporating the fixed effects of lower-level predictors to the random intercept model to quantify the within-person variability. The model was represented by equations as below:

Level 1 equation (modeling within-person variation):

$$DV_{ij} = \beta_{0j} + \beta_{1j}(IV_{L1})_{ij} + \dots + \beta_{nj}(IV_{Ln})_{ij} + e_{ij} \quad (4)$$

Level 2 equations (modeling between-person variation):

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \quad (5)$$

$$\beta_{1j} = \gamma_{10} \quad (6)$$

$$\dots \quad (7)$$

$$\beta_{nj} = \gamma_{n0} \quad (8)$$

where $IV_{L1} \dots IV_{Ln}$ represent lower-level predictors related to trials, such as ToR lead time, *sequence of trial* within participants, and *remaining time to exit*. The *remaining time to exit* was centered at the mean at the within-person level by subtracting a participant’s *average remaining time to exit* of 12 trials from the recorded *remaining time to exit* in each trial. $\beta_{1j} \dots \beta_{nj}$ denote the slope of the predictors for participant *j*; $\gamma_{10} \dots \gamma_{n0}$ denote the

fixed effect (or average slope) of predictors across participants.

In the next step, higher-level predictors were incorporated into the model in Step 2 to quantify the variation in intercepts across participants. Add the participant-level predictors to equation (5) obtained the following equation:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(IV_{H1})_j + \dots + \gamma_{0n}(IV_{Hn})_j + \mu_{0j} \quad (9)$$

where the lone “j” in subscripts represents participants; $IV_{H1} \dots IV_{Hn}$ represent higher-level predictors related to participants, such as driver’s age, annual mileage, initial trust, and *average remaining time to exit*. Specifically, age, annual mileage, and initial trust were grand means centered by subtracting the sample mean from each participant’s data. The initial trust ratings on a 9-point Likert scale were converted to three levels for analysis, namely low (for ratings 1–3), medium (for ratings 4–6), and high (for ratings 7–9). The *average remaining time to exit* was the mean of *remaining time to exit* in 12 trials for each participant. $\gamma_{01} \dots \gamma_{0n}$ denote the fixed effect (or average slope) of the predictors across participants. The Level 1 equation (4) and Level 2 equations (6), (7), (8) remained in the model.

Finally, a likelihood ratio test was performed to determine whether there was significant improvement after incorporating higher-level predictors to the model and then to finalize the model. The detail of the multilevel modeling test and the final model for each dependent measure were described in section 3.1.

The predictors in multilevel modeling tests were summarized in Table 5. The trial-level and participant-level predictors varied across dependent measures due to different data sets and hypotheses for analysis. It was hypothesized that the *remaining time to exit* was an influential factor of driver SA and subjective evaluations in the experiment setting, and therefore was included as a trial-level predictor in the modeling tests. As the exiting outcome was associated with the sign of *remaining time to exit*, namely a positive value of *remaining time to exit* indicating a successful exit and a negative value indicating a failure, the *remaining time to exit* was not considered in the model of exiting outcome. The *remaining time to exit* was also not included in the model of takeover reaction time, as it was obtained by subtracting the ToR lead time by takeover reaction time. Moreover, the exit type was not considered in the models as two types of freeway exits were designed for specific levels of ToR lead time, namely exit only (in 6, 10, 14, 18, 25, 45 s conditions) and not exit only (in 8, 12, 16, 20, 30, 60 s conditions). Including the categorical ToR lead time and exit type in a model would result in redundant parameters.

2.6. Procedure

The experiment was supervised using Zoom, a software-based conference room. After joined the meeting, the participants were asked to show their driver license in the web camera and logged in to the Gorilla.sc platform using a web link and subject ID. To start with, a consent form was presented with detailed information about the study. Only by checking the boxes that identified their will to participate and that

Table 5
Predictors in multilevel modeling tests.

Dependent measure		Data set	Trial-level predictor	Participant-level predictor
Driver SA	Overall/Level 1/Level 2/Level 3	317 trials (excluding trials with exiting failure or video loading delay)	ToR lead time, <i>sequence of trial</i> , <i>remaining time to exit</i>	Age, annual mileage, initial trust, <i>average remaining time to exit</i>
	SA response accuracy			
Subjective evaluations	Readiness for takeover	350 trials (excluding trials with video loading delay)	ToR lead time, <i>sequence of trial</i>	
	Post-trial trust			
	Workload			
Takeover performance	Acceptance			
	Exiting outcome			
	Takeover reaction time			

*Note: Single-level general linear modeling tests were performed for *absolute error of distance* and appropriateness of ToR timing with ToR lead time, *sequence of trial*, and *remaining time to exit* as inputs. See details in section 3.1.5 and 3.3.2, respectively.

Table 6

Descriptive means (standard deviations) of driver SA measures.

ToR lead time (s)	6	8	10	12	14	16	18	20	25	30	45	60
Sample size	27	20	28	28	29	24	29	27	28	26	26	25
Overall SA (%)	63.6 (16.4)	57.1 (16.9)	63.4 (22.8)	69.4 (22.4)	68.7 (24.5)	84.0 (15.5)	77.9 (18.4)	72.2 (21.4)	72.9 (25.8)	75.0 (17.8)	73.7 (24.6)	67.3 (20.1)
Level 1 SA (%)	60.2 (29.6)	47.5 (18.0)	59.8 (31.4)	66.1 (32.8)	71.6 (28.9)	84.4 (19.2)	87.9 (22.8)	77.8 (26.3)	83.0 (27.3)	80.8 (19.1)	82.7 (29.8)	80.0 (17.7)
Level 2 SA (%)	63.9 (23.3)	67.5 (21.6)	71.4 (20.1)	75.0 (22.6)	68.1 (29.0)	84.4 (20.6)	75.0 (25.9)	69.4 (29.7)	71.4 (29.4)	67.3 (29.0)	70.2 (30.8)	56.0 (32.5)
Level 3 SA (%)	66.7 (20.8)	57.5 (29.4)	58.9 (28.2)	67.0 (25.5)	66.4 (27.0)	83.3 (17.5)	70.7 (23.2)	69.4 (22.3)	64.3 (30.0)	76.9 (23.4)	68.3 (25.1)	66.0 (24.9)
Absolute error of distance (ft)	139.3 (112.0)	231.3 (178.8)	154.6 (127.7)	254.3 (221.5)	263.6 (181.6)	213.6 (189.2)	243.5 (219.3)	342.1 (295.2)	330.1 (369.1)	451.6 (449.2)	517.4 (577.1)	1005.3 (1120.0)

*Note: Only the trials with successful exits and no loading delay were included.

requested for consent to video recording, the experiment could start. The participants first filled in a demographic survey including one 9-point Likert scale question that asked them to rate their initial trust in vehicular warning systems based on what they had known or experienced. Then they were given instructions on performing the task in the experiment. Before the formal test, there was a 2-minute practice for participants getting to know how the experiment works and what they needed to do. Participants were also asked to adjust the volume on computer to a comfortable level. While the subjective vehicle was driving in the automation mode, participants were required to play a game named 2048 5×5 on their smart phones for the highest score. As requested by an audio ToR, the participants performed the takeover task when they believed they were able to resume control and exit from the freeway safely. The takeover task required participants to press the spacebar on the laptop or desktop computer to simulate that they disengaged the automation mode and took over control. Following the SAGAT method, when the video stimuli stopped playing as soon as the spacebar was pressed, the participants completed queries to measure their SA and subjective evaluation on the ToR system (see Appendix 1) and started a new trial afterwards.

3. Results

3.1. Relationship between ToR lead time and drive SA

The analysis in section 3.1 included 317 trials. Thirty-five trials with exiting failures were excluded in view of the fact that it was no longer meaningful to evaluate driver SA after they had missed the chance to exit safely and successfully. Ten trials were removed from the analysis in which a blank loading page appeared before the play of video stimuli, which was observed in the Zoom video recordings of participants' shared screens. Although gorilla.sc loads the resources for the future trials in advance of presentation to minimize potential delays due to network speed (Anwyl-Irvine et al., 2020), the MB worth of videos can be difficult for participants with slower internet connects to download, particularly with the increasing distance from the Gorilla server based in

the EU. The trials with video loading delay were excluded as the takeover reaction time recorded by Gorilla were inaccurate. It led to inaccurate values of the variable *remaining time to exit*, which was calculated by ToR lead time subtracting takeover reaction time.

The descriptive statistics in Table 6 presented the sample size as well as the sample means and standard deviations of 5 SA measures at different levels of ToR lead times.

Following Heck et al.'s (2013) "build-up" strategy, the Wald Z tests and intraclass correlation coefficients (ICCs) indicated a substantial clustering of data in the response accuracy of overall SA ($Z = 2.77, p = .006$; ICC = 0.21), Level 1 SA ($Z = 2.72, p = .007$; ICC = 0.21), Level 2 SA ($Z = 2.64, p = .008$; ICC = 0.17), and Level 3 SA ($Z = 2.24, p = .025$; ICC = 0.13), which supported the use of linear mixed modeling with 30 participants as higher-level units. As only a trivial amount of non-

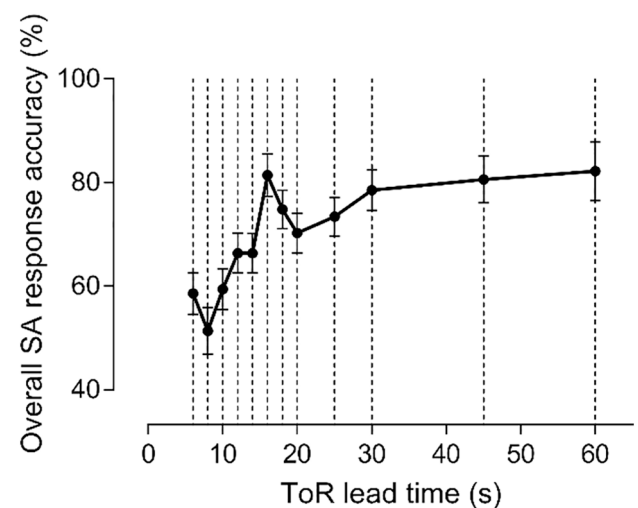


Fig. 4. The relationship between overall SA response accuracy and ToR lead time (EMMs; error bars: ± 1 SD).

Table 7

Mixed model results of effects on overall SA response accuracy.

Effect	F value	Variance	Wald Z	Estimate	Standard Error	df	t value	p
Fixed effects								
Intercept				85.32	6.27	311.97	13.61	<0.001
Remaining time to exit				-0.56	0.16	290.20	-3.58	<0.001
ToR lead time ^a	5.27					288.63		<0.001
Sequence of trial ^a	2.79					288.44		0.002
Random effects								
Intercept		113.06	3.04					0.002
Residual		278.24	11.98					<0.001

Note: a. A categorical variable (12 levels). Estimates and t values for the first 11 levels contrasting the last reference category were omitted.

independence was observed in the *absolute error of distance* at the participant level ($Z = 0.36, p = .72$; $ICC = 0.01$), a single-level general linear model was built with ToR lead time, *sequence of trial*, and *remaining time to exit* as inputs. A significance level of 0.025 for one-tailed Wald Z tests and a threshold of 0.05 for ICCs were used for all of the multilevel modeling tests in present study based on the criteria proposed in Heck et al. (2013).

Next, models were tested by adding the fixed effects of lower-level predictors to the random intercept model. Three trial-level predictors, i.e., ToR lead time, *sequence of trial* within participants, and *remaining time to exit*, were included in the model to quantify the within-person variability. Two types of freeway exits (i.e., exit only and not exit only) were created to increase the variation in takeover scenarios. It was assumed that driver SA was not significantly affected by the exit type. To verify the assumption, one-way ANOVA tests were conducted and results showed that there were not significant effects of exit type on overall SA ($F(1, 315) = 0.23, p = .63, \eta_p^2 = 0.001$), Level 1 SA ($F(1, 315) = 0.06, p = .81, \eta_p^2 < 0.001$), Level 2 SA ($F(1, 315) < 0.001, p = .98, \eta_p^2 < 0.001$), or Level 3 SA ($F(1, 315) = 2.46, p = .12, \eta_p^2 = 0.01$).

The multilevel modeling test on the effects of participant-level predictors on each SA response accuracy measure and the single-level general linear modeling test on *absolute error of distance* were reported in the following sections in 3.1.

3.1.1. Overall SA response accuracy

The multilevel modeling test did not find any significant effects of participant-level variables (i.e., driver's age, annual mileage, initial trust, and *average remaining time to exit*) on overall SA. The likelihood ratio test also showed that there was no significant improvement by incorporating higher-level predictors to the model, $\chi^2(4) = 6.93, p = .14$. Therefore, the final mixed model of overall SA was obtained with ToR lead time, *sequence of trial*, and *remaining time to exit* as predictors.

The results revealed a significant effect of ToR lead time on the overall SA response accuracy at a significance level of 0.05 (which was applied for the following F tests), $F(11, 288.63) = 5.27, p < .001$. The results of linear mixed model were presented in Table 7. In general, providing a longer ToR lead time significantly improved the overall driver SA for resuming manual control to exit a freeway.

Given the between- and within-person variability and the unbalanced sample sizes for 12 groups of ToR lead time, the estimated marginal means (EMMs) adjusted for any other variables in the model were plotted as Fig. 4 as well as the following trend graphs (Figs. 5–15). The trend graph in Fig. 4 showed that the overall SA response accuracy was higher in 16–60 s trials than in 6–14 s trials. There was a sharp increase

and a subsequent drop in the overall SA at 14–20 s with the peak at 16 s. When the ToR had a lead time longer than 16 s, no significant increase was found in the overall SA. The pairwise comparisons showed that the overall SA was significantly higher at 60 s, 45 s, and 30 s than 6–14 s; higher at 25 s, 20 s, and 18 s than 6–10 s; higher at 16 s than 6–14 s, and 20 s; higher at 14 s and 12 s than 8 s. Based on the trend graph and pairwise comparisons, the increase in overall SA with the ToR lead time leveled off at 16–30 s.

The significant effect of *sequence of trial* on the overall SA response accuracy was also observed, $F(11, 288.44) = 2.79, p = .002$. The pairwise comparison test showed that the first two trials within individuals resulted in significantly lower overall SA than the subsequent trials in general. The *remaining time to exit* also significantly affected the overall SA, $b = -0.56, t(290.20) = -3.58, p < .001, 95\% \text{ CI } [-0.86, -0.25]$. The overall SA response accuracy was approximately 0.6% higher when the *remaining time to exit* was 1 s shorter, namely when the subject vehicle was 88 feet (at 60 mph) or 73 feet (at 50 mph) closer to the freeway exit.

3.1.2. Level 1 SA response accuracy

The multilevel modeling test did not reveal any significant effects of participant-level predictors on Level 1 SA. The likelihood ratio test also did not find any significant improvement by incorporating higher-level predictors to the model, $\chi^2(4) = 4.27, p = .37$. Therefore, the final mixed model of Level 1 SA included ToR lead time, *sequence of trial*, and *remaining time to exit* as predictors.

The ToR lead time was a significant predictor of Level 1 SA response accuracy, $F(11, 287.80) = 6.80, p < .001$. The trend graph in Fig. 5 showed that the Level 1 SA significantly improved with the increase of ToR lead time in general. There was a dramatic increase in Level 1 SA when the ToR lead time was between 6 s and 18 s, and the increase tended to be gentle for trials with a ToR lead time longer than 18 s. Results of pairwise comparisons showed that Level 1 SA was significantly higher at 60 s, 45 s, and 18 s than 6–14 s; higher at 30 s, 25 s, 20 s, and 16 s than 6–12 s; higher at 14 s than 6–10 s; higher at 12 s, 10 s, and 6 s than 8 s. Based on the trend graph and pairwise comparisons, the increase in Level 1 SA with the ToR lead time leveled off at 18–45 s.

The *sequence of trial* also significantly affected Level 1 SA response accuracy, $F(11, 287.61) = 2.34, p = .01$. The pairwise comparisons showed that the Level 1 SA was significantly lower in the 1st trial than in the 3rd, 5th, 6th, 9th and subsequent trials; lower in the 2nd trial than in the 5th, 6th, 9th, 10th, and 11th trials; lower in the 4th trial than in the 5th, 6th, and 9th trials; lower in the 7th trial than in the 9th trial. The *remaining time to exit* was not a significant factor in predicting Level 1 SA

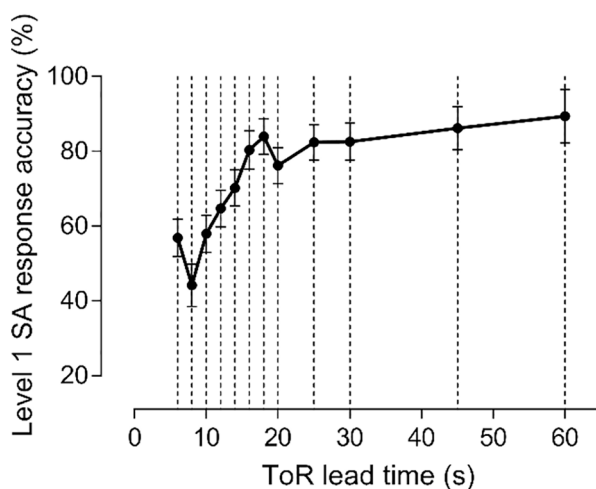


Fig. 5. The relationship between Level 1 SA response accuracy and ToR lead time (EMMs; error bars: ± 1 SD).

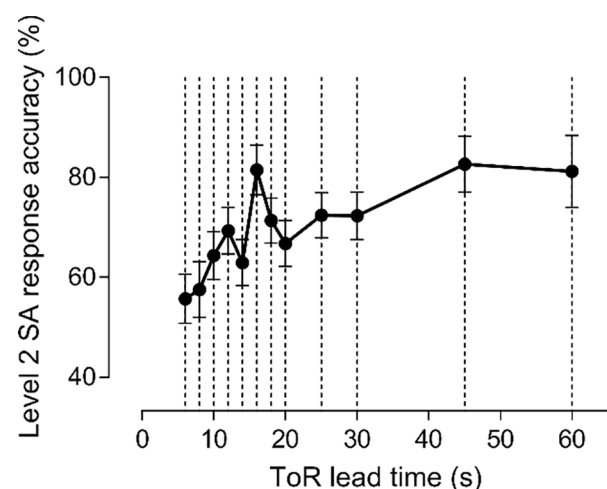


Fig. 6. The relationship between Level 2 SA response accuracy and ToR lead time (EMMs; error bars: ± 1 SD).

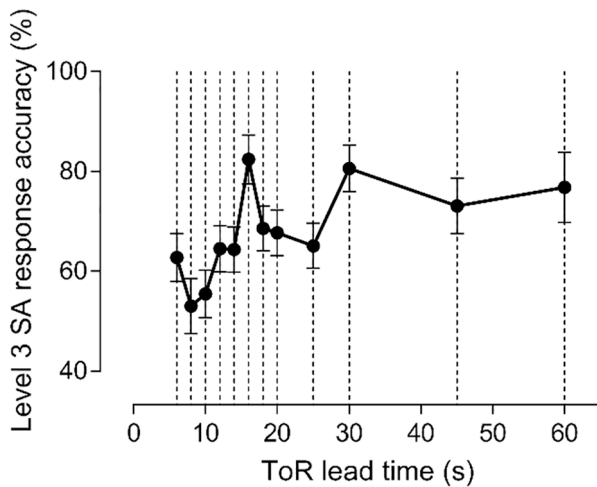


Fig. 7. The relationship between Level 3 SA response accuracy and ToR lead time (EMMs; error bars: ± 1 SD).

response accuracy, $b = -0.32$, $t(289.34) = -1.67$, $p = .10$, 95% CI $[-0.71, 0.06]$.

3.1.3. Level 2 SA response accuracy

The multilevel modeling test revealed a significant effect of the participant-level variable *average remaining time to exit* on Level 2 SA response accuracy, $b = -1.19$, $t(31.04) = -3.17$, $p = .003$, 95% CI $[-1.96, -0.43]$; whereas, no significant effects of driver's age, annual mileage, or initial trust were found on Level 2 SA. The likelihood ratio test showed that there was significant improvement by incorporating higher-level predictors to the model, $\chi^2(4) = 12.03$, $p = .02$. Therefore, the mixed model of Level 2 SA was finalized with lower-level predictors (i.e., ToR lead time, *sequence of trial*, and *remaining time to exit*) and higher-level predictors (i.e., driver's age, annual mileage, initial trust, and *average remaining time to exit*).

The ToR lead time was found to be significantly associated with Level 2 SA response accuracy, $F(11, 290.10) = 2.40$, $p = .01$. Fig. 6 showed a general increase in Level 2 SA along with the ToR lead time. The pairwise comparison test revealed a positive effect of ToR lead time on Level 2 SA. Specifically, Level 2 SA was significantly higher at 60 s than 6 s, 8 s, and 14 s; higher at 45 s and 16 s than 6–10 s, 14 s, and 20 s; higher at 30 s, 25 s, and 18 s than 6 s and 8 s; higher at 12 s than 6 s. No apparent saturation was observed from the trend graph or comparison tests.

Results showed that the *remaining time to exit* was significantly associated with Level 2 SA, $b = -0.95$, $t(291.48) = -4.63$, $p < .001$,

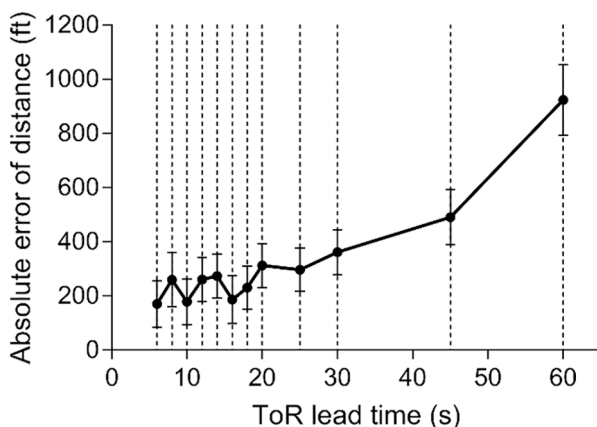


Fig. 8. The relationship between *absolute error of distance* and ToR lead time (EMMs; error bars: ± 1 SD).

95% CI $[-1.35, -0.54]$, which indicated that the Level 2 SA response accuracy was approximated 1% higher when participants left 1 s shorter (namely 88 feet (at 60 mph) or 73 feet (at 50 mph) closer) to exit the freeway. The *sequence of trial* did not significantly influence Level 2 SA, $F(11, 290.09) = 1.68$, $p = .08$.

3.1.4. Level 3 SA response accuracy

The multilevel modeling test revealed a significant effect of annual mileage on Level 3 SA, $b = -0.0009$, $t(29.74) = -2.17$, $p = .04$, 95% CI $[-0.002, -0.00005]$. However, the likelihood ratio test showed that there was no significant improvement by incorporating higher-level predictors to the model, $\chi^2(4) = 7.43$, $p = .11$. Therefore, the mixed model of Level 3 SA was obtained with ToR lead time, *sequence of trial*, and *remaining time to exit* as predictors.

The ToR lead time had a significant effect on Level 3 SA response accuracy, $F(11, 288.26) = 3.20$, $p < .001$. The Fig. 7 did not show any obvious trend in Level 3 SA along with the ToR lead time. The pairwise comparison test showed that Level 3 SA was significant higher at 60 s, 45 s, 20 s, and 18 s than 8 s and 10 s; higher at 30 s and 16 s than 6–14 s and 18–25 s. No apparent saturation was observed from the trend graph or comparison tests.

The *sequence of trial* significantly affected Level 3 SA response accuracy, $F(1, 287.96) = 3.02$, $p = .001$. The pairwise comparison test indicated that Level 3 SA was significantly lower in the 1st trial than in the 3rd, 5th, 6th, 7th, 8th, 9th, 11th, and 12th trials; lower in the 2nd trial than in the 3rd, 6th, 7th, 8th, 9th, 11th, and 12th; lower in the 4th trial than in the 3rd trial. The *remaining time to exit* also had a significant effect on Level 3 SA, $b = -0.43$, $t(290.55) = -2.14$, $p = .03$, 95% CI $[-0.82, -0.03]$. Participants had 0.4% higher Level 3 SA response accuracy when participants left 1 s shorter (namely 88 feet (at 60 mph) or 73 feet (at 50 mph) closer) to exit the freeway.

3.1.5. Absolute error of distance

The general linear modeling test showed that ToR lead time was significantly associated with *absolute error of distance*, $F(11, 293) = 2.14$, $p = .02$, $\eta_p^2 = 0.07$. As shown in Fig. 8, the increase in *absolute error of distance* had a stable slope of low gradient for the ToR lead time of 6–30 s, and it became relatively sharp at 30–60 s. The pairwise comparison test showed that the *absolute error of distance* was significantly greater at 60 s than any other levels; greater at 45 s than 6 s, 10 s, and 16 s. The trend graph and multiple pairwise comparisons revealed that drivers' awareness of remaining travel distance to exit dramatically reduced at longer ToR lead times ranging from 30 s to 45 s. No significant association was found between *absolute error of distance* and the *sequence of trial* ($F(11, 293) = 0.97$, $p = .48$, $\eta_p^2 = 0.04$) or *remaining time to exit* ($F(1, 293) = 1.36$, $p = .25$, $\eta_p^2 = 0.01$).

3.2. Effects of ToR lead time on takeover performance

The analysis and results of the effects of ToR lead time on two takeover performance measures, i.e., exiting outcome and takeover reaction time, were reported in this section. The analysis included 350 trials after removing 10 trials with video loading delay. The Table 8 presented the sample size, the successful exiting rate as well as the sample mean and standard deviation of takeover reaction time for each level of ToR lead time.

3.2.1. Exiting outcome

The multilevel logistic regression was conducted to examine the effects of ToR lead time on the binary exiting outcome. Following the general framework of multilevel modeling by Heck et al. (2013), the null model with intercept only was tested first to determine whether there was significant non-independence within participants on the exiting outcome. Results revealed significant between-person variability at a significance level of 0.025 for one-tailed Wald Z test ($Z = 2.55$, $p = .01$) and it yielded an ICC of 0.46, which indicated a substantial clustering

Table 8

Descriptive means (standard deviations) of takeover performance measures.

ToR lead time (s)	6	8	10	12	14	16	18	20	25	30	45	60
Sample size	30	30	29	29	30	29	30	29	29	28	29	28
Successful exiting rate (%)	90	67	97	93	97	83	97	90	97	93	90	90
Takeover reaction time (s)	4.48 (1.39)	7.02 (2.12)	7.15 (2.15)	8.21 (3.24)	9.89 (3.33)	11.04 (4.39)	11.55 (4.49)	12.75 (5.47)	15.08 (6.71)	17.01 (9.62)	21.97 (15.66)	26.82 (23.26)

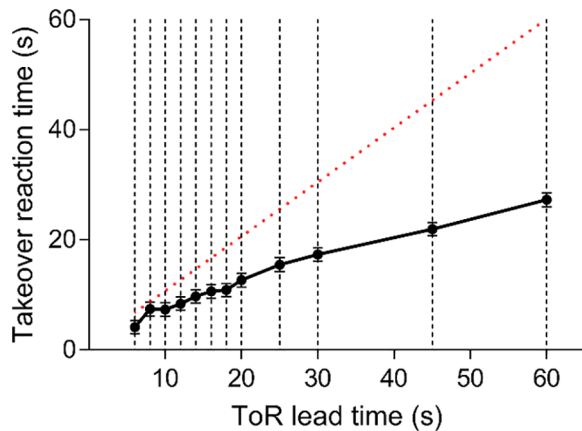


Fig. 9. The relationship between takeover reaction time and ToR lead time (EMMs; error bars: ± 1 SD). *Note: The dash line in red denotes the ToR lead time given, namely the available time for takeover. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

effect in the data at a threshold of 0.05 (Heck et al., 2013). The lower-level predictors were ToR lead time and sequence of trial. Adding higher-level predictors (i.e., driver's age, annual mileage, initial trust, and average remaining time to exit) significantly improved the model ($\chi^2(4) = 100.48$, $p < .001$) and therefore were included in the final model. Results indicated that the exiting outcome was not significantly affected by ToR lead time ($F(1, 333) = 0.75$, $p = .39$) or the sequence of trial ($F(11, 333) = 0.89$, $p = .55$). No significant effects of age, annual mileage, or initial trust were observed on exiting outcome. Nevertheless, the average remaining time to exit was a significant predictor of exiting outcome such that individuals who tended to leave longer manual driving time before exiting were more likely to exit a freeway successfully, $b = 0.19$, $t(333) = 2.83$, $p = .005$, 95% CI [0.06, 0.32].

3.2.2. Takeover reaction time

The trend graph in Fig. 9 presented a linear increase in takeover reaction time along with the ToR lead time, with the dash line in red denoting the ToR lead time as the maximal available time for takeover. Following the general “build-up” strategy for model testing used by Heck et al. (2013), the results of testing the random intercept model

revealed a substantial clustering in data, $Z = 3.01$, $p = .003$; ICC = 0.23, which supported the use of the linear mixed modeling technique with 30 participants as higher-level units. The lower-level predictors in the model were ToR lead time and *sequence of trial*. Adding higher-level variables significantly improved the model ($\chi^2(4) = 91.54$, $p < .001$) and therefore were included in the final mixed model. Results found a significant association between ToR lead time and takeover reaction time, $b = 0.41$, $t(350) = 17.31$, $p < .001$, 95% CI [0.36, 0.46], which indicated that participants' takeover reaction time was 0.4 s longer on average when they were given 1 s longer lead time to take over. The *average remaining time to exit* was significantly associated with takeover reaction time, $b = -0.96$, $t(350) = -16.03$, $p < .001$, 95% CI [-1.08, -0.85]. No significant effects of drivers' age, annual mileage, initial trust, or the *sequence of trial* on takeover reaction time were observed.

3.3. Effects of ToR lead time on subjective evaluations

Considering the individual differences in subjective experience, Heck et al.'s (2013) “build-up” strategy was used for multilevel modeling test for each measure of subjective evaluations, namely readiness for takeover, post-trial trust, workload, and acceptance. The analysis in section 3.3 were performed based on the same data sets in section 3.1, namely

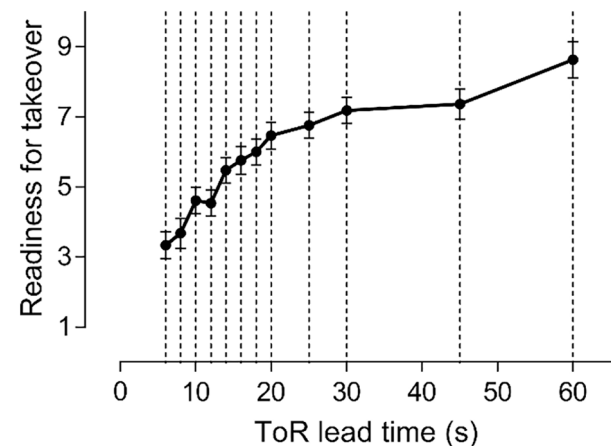


Fig. 10. The relationship between readiness for takeover and ToR lead time (EMMs; error bars: ± 1 SD).

Table 9

Descriptive means (standard deviations) of subjective evaluation measures on 9-point Likert scales.

ToR lead time	6 s	8 s	10 s	12 s	14 s	16 s	18 s	20 s	25 s	30 s	45 s	60 s
Sample size	27	20	28	28	29	24	29	27	28	26	26	25
Readiness for takeover	3.6 (2.7)	4.0 (2.1)	5.0 (2.4)	5.0 (2.3)	5.7 (2.1)	6.1 (2.0)	6.2 (2.0)	6.6 (1.9)	6.7 (2.0)	7.1 (1.7)	6.9 (1.9)	7.3 (1.6)
Appropriateness of ToR timing	7.3 (2.1)	6.4 (1.9)	6.2 (1.9)	5.2 (1.7)	5.2 (1.5)	5.1 (1.6)	4.7 (1.2)	4.1 (1.3)	4.0 (1.6)	3.3 (1.4)	3.3 (1.7)	2.9 (1.7)
Post-trial trust	4.0 (2.5)	4.8 (2.0)	4.9 (2.0)	5.8 (1.9)	5.6 (1.8)	6.1 (1.8)	6.4 (1.8)	6.1 (1.7)	6.4 (1.7)	6.3 (1.9)	6.5 (2.0)	6.4 (1.8)
Workload	5.3 (2.3)	4.9 (2.1)	4.9 (2.1)	4.3 (1.9)	4.4 (2.0)	4.2 (1.7)	4.3 (2.1)	3.7 (2.0)	3.8 (1.9)	3.5 (1.7)	3.6 (1.4)	3.6 (2.0)
Acceptance	4.5 (2.2)	4.9 (1.6)	5.3 (1.6)	5.6 (1.8)	5.6 (1.5)	6.2 (1.3)	6.0 (1.4)	6.3 (1.7)	6.1 (1.4)	6.5 (1.5)	6.5 (1.4)	6.5 (1.8)

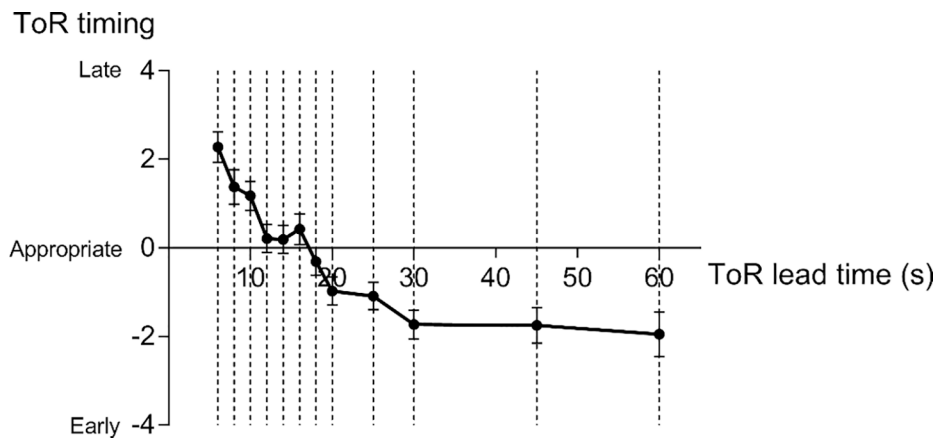


Fig. 11. The relationship between appropriateness of ToR timing and ToR lead time (EMMs; error bars: ± 1 SD).

317 trials after excluding the trials with exiting failures in view of the fact that it was no longer meaningful to evaluate the ToR system when participants failed to exit the freeway. The trials when video loading delay occurred were also removed from analyses due to the inaccurate values of *remaining time to exit*. The descriptive statistics in Table 9 presented the sample size, sample mean, and standard deviation of each subjective evaluation measure for 12 levels of ToR lead time.

3.3.1. Readiness for takeover

The random intercept model testing yielded a substantial clustering in the readiness rating data ($Z = 3.18$, $p = .001$; ICC = 0.32), which supported the multilevel modeling test. The inputs of the final mixed model were ToR lead time, *sequence of trial*, and *remaining time to exit* after examining the insignificant improvement of considering higher-level predictors in the model ($\chi^2(4) = 4.91$, $p = .30$). It was found that the ToR lead time significantly affected participants' readiness for takeover, $F(11, 288.94) = 14.52$, $p < .001$. The relationship between ToR lead time and drivers' readiness for takeover was presented as Fig. 10. The pairwise comparison test showed that participants had more readiness at 60 s than any other levels; more readiness at 45 s, 30 s, and 25 s than 6–18 s; more readiness at 20 s than 6–14 s; more readiness at 18 s, 16 s, and 14 s than 6–12 s; more readiness at 12 s and 10 s than 6 s and 8 s. From Fig. 10, the readiness increased with ToR lead time at 6–25 s, stabilized at 25–45 s, and then increased at 45–60 s. The trend graph and pairwise comparisons did not suggest any saturation of the positive effects of ToR lead time on participants' readiness for takeover, which did not meet the expected trend in Table 4.

3.3.2. Appropriateness of ToR timing

The random intercept model testing did not show any statistically significant variance at the participant level, $Z = 1.14$, $p = .25$. The ICC (=0.04) also indicated a trivial amount of non-independence in the appropriateness at the trial's level. Thus, a single-level general linear model was tested with ToR lead time, *sequence of trial*, and *remaining time to exit* as inputs. Results showed that the ToR lead time significantly affected participants' evaluations on the appropriateness of timing to issue ToR, $F(11, 293) = 9.97$, $p < .001$, $\eta_p^2 = 0.27$. Fig. 11 presented the relationship between ToR lead time and the appropriateness of ToR timing, which was point 5 centered by subtracting 5 from original ratings (i.e., 1 = extremely early, 9 = extremely late, and 5 = appropriate timing). The pairwise comparison test showed that the perceived appropriateness of the timing to issue ToR was significantly later at 6 s than 10–60 s; later at 8 s and 10 s than 12 s, 14 s, and 18–60 s; earlier at 60 s, 45 s, and 30 s than 6–18 s; earlier at 25 s and 20 s than 6–16 s. The trend in Fig. 11 met the hypothesis in Table 4.

3.3.3. Post-trial trust

The random intercept model testing indicated a substantial clustering effect in participants' post-trial trust, $Z = 3.53$, $p < .001$; ICC = 0.53. A mixed model was run with ToR lead time, *sequence of trial*, *remaining time to exit* as predictors after examining the insignificant improvement of incorporating higher-level predictors to the model ($\chi^2(4) = 4.87$, $p = .30$). Results showed that the ToR lead time significantly affected participants' post-trial trust in the automation system, $F(11, 288.50) = 13.39$, $p < .001$. The relationship between ToR lead time

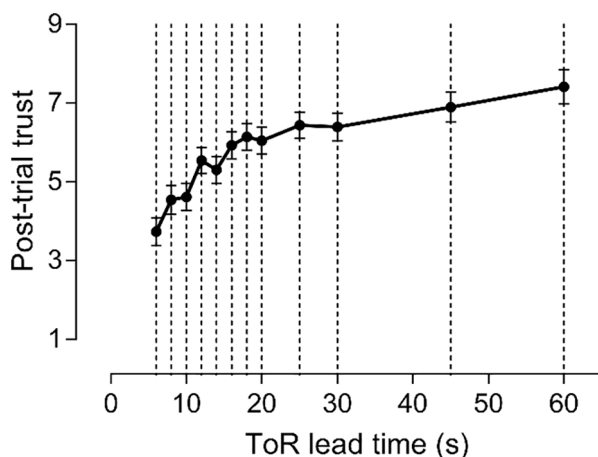


Fig. 12. The relationship between post-trial trust and ToR lead time (EMMs; error bars: ± 1 SD).

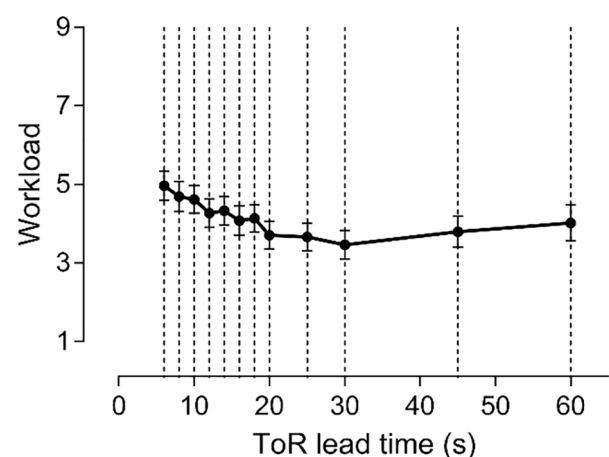


Fig. 13. The relationship between workload and ToR lead time (EMMs; error bars: ± 1 SD).

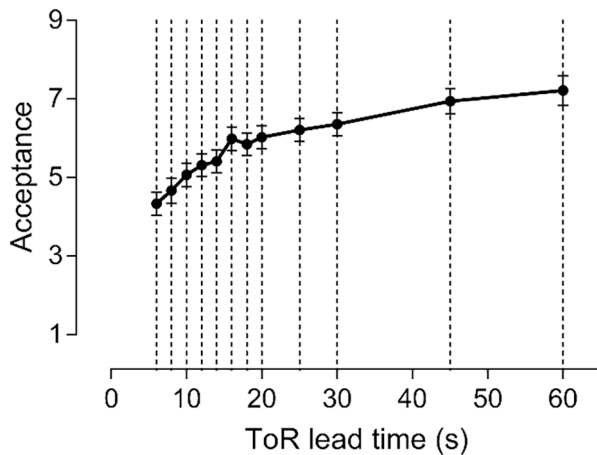


Fig. 14. The relationship between acceptance and ToR lead time (EMMs; error bars: ± 1 SD).

and post-trial trust was presented as Fig. 12. The pairwise comparison test showed that the post-trial trust was significantly higher at 60 s than 6–30 s; higher at 45 s than 6–20 s; higher at 30 s, 25 s, and 18 s than 6–14 s; higher at 20 s and 16 s than 6–10 s and 14 s; higher at 14 s and 12 s than 6–10 s; higher at 10 s and 8 s than 6 s. From Fig. 12, driver trust increased with ToR lead time at 6–18 s, stabilized at 18–30 s, and then increased again at 30–60 s. The trend graph and pairwise comparisons did not suggest any saturation of the positive effects of ToR lead time on participants' post-trial trust in the ToR system, which did not meet the expected trend in Table 4.

3.3.4. Workload

The random intercept model testing yielded a substantial clustering in participants' workload, $Z = 3.54$, $p < .001$; ICC = 0.55. The final mixed model incorporated the ToR lead time, *sequence of trial*, and *remaining time to exit* as predictors after examining the insignificant improvement of incorporating higher-level predictors to the model ($\chi^2(4) = 4.71$, $p = .32$). Results showed that the ToR lead time was a significant factor of workload, $F(11, 288.64) = 3.14$, $p = .001$. The relationship between ToR lead time and workload was presented in Fig. 13. The pairwise comparison test showed that the workload was significantly lower at 45 s and 20 s than 6–10 s; lower at 30 s than 6–14 s and 18 s; lower at 25 s than 6–10 s and 14 s; lower at 18 s, 16 s, 14 s, and 12 s than 6 s. Based on the trend graph and pairwise comparisons, the positive effects of longer ToR lead time on reducing drivers' workload reached the maximum at 20–45 s, which did not meet the expected trend in Table 4.

3.3.5. Acceptance

A substantial clustering in participants' acceptance was found from the random intercept model testing, $Z = 3.51$, $p < .001$; ICC = 0.51. The final mixed model included ToR lead time, *sequence of trial*, and *remaining time to exit* as inputs after examining the insignificant improvement of considering higher-level predictors in the model ($\chi^2(4) = 6.63$, $p = .16$). It was found that participants' acceptance to the ToR system significantly depended on the ToR lead time, $F(11, 288.57) = 9.33$, $p < .001$. Fig. 14 presented the relationship between ToR lead time and. The pairwise comparison test found that the ToR lead time of 60 s and 45 s were more accepted than 6–30 s; 20–30 s and 16 s were more accepted than 6–14 s; 18 s was more accepted than 6–12 s; 14 s and 12 s were more accepted than 6 s and 8 s; 10 s was more accepted than 6 s. The trend graph and pairwise comparisons demonstrated that the increase in drivers' acceptance to the warning system with the ToR lead time leveled off at 16–45 s, which met the expected trend of acceptance in Table 4.

4. Discussion

The present study mainly aimed to investigate the relationship between ToR lead time and driver SA when executing takeover to exit from freeways in conditionally automated driving. In the meanwhile, this study explores the effect of ToR lead time on drivers' takeover reaction time and subjective evaluations on the ToR warning system. The following sections discuss the hypothesis testing on the relationship between driver SA and takeover reaction time under the varying level of ToR lead time, make recommendations on designing the ToR lead time for freeway exiting, and provide insights on enhancing driver SA for resuming manual control from automated vehicles that are incapable of maneuvering itself to exit from freeways.

4.1. Interpretation of hypotheses testing results

• Driver SA and ToR lead time

The trend graphs of overall SA response accuracy (see Fig. 4) and takeover reaction time (see Fig. 9) are merged into one figure as shown in Fig. 15. In general, the trend line of overall SA meets our hypothesis on area ② and ③ in Fig. 1. Firstly, there exists a turning point t_{SA} at 16 s lead time for driver SA. The slope of increase in overall SA is steep at 8–16 s, which verifies the dramatic improvement in driver SA before the lead time is long enough for drivers to gain sufficient SA (corresponding to area ② in Fig. 1). Although the trend graph shows a decrease at 18 s and 20 s, the comparison test reveals no significant change in the overall SA for the trials with a ToR lead time longer than 16 s. Secondly, when the time budget is enough, the slope of SA becomes gentle at 20–60 s, which demonstrates the saturated effect of ToR lead times longer than t_{SA} on driver SA (corresponding to area ③ in Fig. 1). Surprisingly, the trend graph does not reveal a drop in the overall SA when the ToR lead

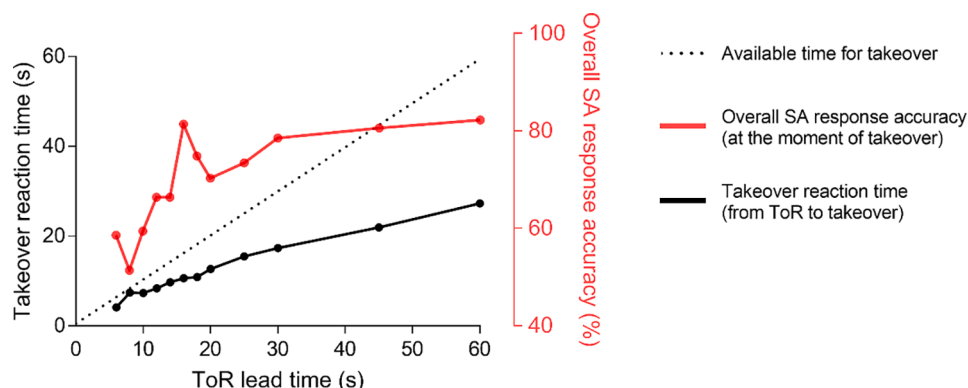


Fig. 15. Merged trend lines of takeover reaction time and overall SA response accuracy with the increase of ToR lead time.

time is at 60 s. It suggests future automated vehicle systems should provide drivers more time if possible to reach enough driver SA for takeover and manually exit from freeways.

The results found that the positive effect of longer ToR lead times on overall SA saturates at 16–30 s, which is longer as compared to 7–12 s reported in Lu et al. (2017). Such difference could be resulted from the different task demand and different highway scenarios. Lu et al. (2017) asked participants to passively watch video clips about freeway driving till the end without the need to take over the AVs, and to reproduce the traffic layout as to the number, positions, and speed of surrounding cars. In comparison, the present study required participants to disengage from a non-driving activity, attend to and comprehend multiple elements (i.e., exit type, movement of two cars ahead, subjective vehicle's signal lights and speed, and speed limit to exit), and finally take over control in a self-paced manner. The higher task demand by the present study could be the reason of longer lead time required by drivers to gain sufficient SA for resuming control.

• Takeover reaction time and ToR lead time

The trend line of takeover reaction time partly aligns with our hypothesis. As shown in Fig. 15, there is an approximate turning point of takeover reaction time at 16 s, which confirms our hypothesis as to the decrease of slope at t_{SA} in Fig. 1. The trend line of takeover reaction time for ToR lead times between 6 s and 8 s is in line with hypothesis in that, the takeover reaction time is approximate to the given lead time and the slope of increase is about 1. It indicates that the average driver tended to take over control at the last moment before exiting from freeways, possibly to make the most use of the time to gain more SA. The “last-moment takeover” was hypothesized to occur for the whole area ② (namely 6–16 s in the present study) when the time budget is not enough for the average driver acquiring good SA. However, for ToR lead times between 8 s and 16 s, the trend line of takeover reaction time in the results does not fully align with the hypothesis. Although the takeover reaction time increases along with the increased ToR lead times, the slope of takeover reaction time is not as great as 1, suggesting that not all drivers take the “last-moment takeover” strategy to get better SA when it is possible. There exists a portion of drivers who took over control of AVs without having sufficient SA and therefore might not be ready to steadily control the vehicles afterwards, which results in a greater risk of driving instability, violation of traffic rules, and even accidents after takeover. The finding highlights the significance of accommodating the variance in driver takeover strategies into the future design of automation driving systems for safe control transitions. Some SA assistance and assessment technologies can be considered to enable the automation system to hand over control to the human driver only when the driver obtains adequate SA. Lastly, the trend of takeover reaction time for ToR lead times longer than 16 s does not provide evidence against the hypothesis for area ③ in Fig. 1. The increase in takeover reaction time with a slope smaller than 1 indicates that, when the time budget is enough for gaining SA, the average driver postponed the takeover action further when given a longer lead time, but did not wait till the last moment to exit from freeways.

Drivers' takeover reaction time for non-critical freeway exiting scenarios was found longer than that for time-critical control transitions reported in the literature (Feldhütter et al., 2017; Gold et al., 2016; Kerschbaum et al., 2015; Körber et al., 2016; Louw et al., 2015; Lorenz et al., 2014; Melcher et al., 2015; Radlmayr et al., 2014). The finding is in line with Eriksson and Stanton (2017), which concluded that drivers took longer to resume control when under no time pressure. Nevertheless, while Eriksson and Stanton (2017) focused on the automation failure in which drivers were prompted to resume control with no time limits, the present study adds value to the analysis of drivers' takeover behavior under a wide range with multiple levels of ToR lead time in freeway exiting scenarios. Moreover, the present study revealed longer takeover reaction time under the same level of ToR lead time comparable to that in existing studies that investigated drivers' takeover

behavior in non-critical control transitions involving freeway exiting scenarios (see Table 1). One possible reason could be that some previous studies included various takeover events (e.g., construction sites, breakdown vehicles) other than freeway exiting, which required drivers' quick reaction and thus shortened the average takeover reaction time. It is also possibly because the freeway exiting scenarios in some studies required lane changing that made drivers feel an urgent need to resume control, whereas the present study did not. The scenario design in the present study was to ensure no bias was introduced by the extra time required for manual lane change. Additionally, no actual manual driving maneuvers were required after takeovers in the web-based study, and it therefore may diminish drivers' sense of urgency for resuming manual control of the vehicle. The low-fidelity experiment also reduces the harm of missing a freeway exit. These causes may lead to drivers' takeover latency in this study.

4.2. Recommended ToR lead time for exiting a freeway

The present study suggested that the ToR lead time had a positive effect on the overall driver SA, and the positive effect saturated at the lead time ranging between 16 s and 30 s. When dived into three SA levels, the same conclusion was made for Level 1 SA that participants' perception of environmental information leveled off when the ToR lead time was between 18 s and 45 s. However, the results of Level 2 and Level 3 SA did not show such apparent saturation of effects. According to the comparison tests, the response accuracy appeared to approach the peak level at 16 s and 45 s for drivers' comprehension of information (Level 2 SA) and at 16 s and 30 s for drivers' projection of operation after takeover (Level 3 SA). Based on these findings, it was concluded that the ToR lead time of 16–60 s was recommended for good overall driver SA and drivers' perception of necessary information for exiting from freeways. Moreover, Level 2 and Level 3 SA was not as good as Level 1 SA in the trials with a ToR lead time longer than 16 s. To help drivers improve their SA at the comprehension and projection levels, additional system assistance and operational-level advice are suggested to be offered for better understanding the situation and required operations after resuming manual control.

The difference between drivers' predicted distance and the actual distance away from the freeway exit increased with the ToR lead time, which indicated a decline in drivers' awareness of remaining travel distance to exit. The results revealed a substantial increase in *absolute error of distance* between 30 s and 60 s, and no significant difference was found between any pairs of ToR lead time between 6 s and 30 s. Therefore, the ToR lead time of 6–30 s was recommended for drivers' better awareness of remaining distance to exit. Additionally, a dynamic visualization of remaining distance to exit is suggested to be provided to drivers for executing takeover actions appropriately and improving the driving performance after takeover (Holländer and Pfleging, 2018).

The results of subjective evaluations yielded diverse results in multiple measures. With regard to the timing of ToRs, the ToR lead time of 12–18 s was rated as the most appropriate, namely neither too late nor too early. Participants' readiness for takeover and post-trial trust in the ToR system increased with the ToR lead time, but did not level off. A ToR lead time of 25–60 s and 18–60 s was recommended for drivers' better readiness and higher trust, respectively. The positive effects of ToR lead time on reducing participants' workload also did not saturate but reached to the maximum at 25–45 s. Participants tended to accept the automation system with a longer ToR lead time and the acceptance leveled off at the lead time between 45 s and 60 s. Our recommendations for the design of ToR lead time for freeway exiting takeovers were given based on the combination of findings. In the current settings of Level 3 automation systems, the ToR lead time of 16–30 s was suggested for drivers gaining the best SA. On this basis, after combining the results of takeover readiness, workload, and trust, the recommended range of ToR lead time would be narrowed down to 25–30 s. However, a ToR lead time of 25–30 s was early for drivers and was less accepted than longer

lead times; Additionally, we suggested some SA recovering assistance at the operational level to be offered to the drivers for helping rebuild SA quickly, reduce workload, and let drivers get more readiness when execute the takeover action.

4.3. Benefits of driver training on using automated driving systems

The *sequence of trial* within each subject was found to be a significant factor to the overall SA as well as Level 1 and Level 3 SA. Generally, drivers had higher SA response accuracy in the later trials as compared to the first few trials. It is noteworthy that the ToR lead time levels are randomized within subjects to minimize the carry-over effect caused by the sequence of trials; the analyses have included the sequence number of trial in the models so that the sequence of trials does not influence the results about driver SA. As the same SA queries were used for all of trials, there was a possibility of carry-over effects that the participants intentionally paid attention to the environmental elements being queried after experiencing a few of trials and getting familiar to the task. The finding implies the potential benefits of offering drivers training on gaining SA for specific takeover situations while using automated driving systems. Although drivers will not be motivated to answer queries in actual use on road, such driver training is expected to foster the consciousness of acquiring necessary information before executing the takeover action for driving safety. Moreover, the results in the present study were obtained based on a single-visit experiment. It is unknown how driver SA and takeover behavior would change with their gaining experience on such automated driving systems on a long-term basis. Future studies could consider the longitudinal design to explore the change over time as to driver SA and takeover performance while using automated driving systems.

4.4. Limitations and future directions

As in-person laboratory experiments were not permitted due to a public health issue in the United States, the web-based experiment was conducted online as a preliminary study. Several limitations need to be considered in the present work. Firstly, the situation awareness was evaluated using the SAGAT method in which the freeze probe was initiated by the participant pressing the spacebar. Although SAGAT is found to be a highly sensitive, reliable, and predictive measure of SA, there is criticism of its deterministic in that SA is pre-defined and measured by comparing against a normative ideal and it is hardly achievable in naturalistic settings (Endsley, 2019; Salmon et al., 2012). Secondly, the post-takeover driving performance cannot be collected as implied driver SA. As the simulation stopped once participants pressed the spacebar, it is unknown whether participants were able to exit from the freeway successfully after takeover. Future studies in the laboratory will combine SAGAT with the post-takeover performance and eye movement measures for more compelling results. Thirdly, the browser-based experiment was conducted over the internet, which needed participants downloading videos and audios on their ends. Although a video clip rather than a complete video was used in each trial, there were still 10 trials with video loading delay caused by poor internet connection, which resulted in inaccurate values of takeover reaction time recorded by Gorilla.sc; therefore, these trials were removed from the statistical analysis. The different sound volume on participants' computers could result in different loudness levels and therefore affect their takeover reaction times to the audio stimulus. Fourthly, single-item scales have limitations in measuring the multifaceted notions such as trust and workload. Future studies that focus on the subjective evaluations are recommended to use multi-item questionnaires that have been

validated. Lastly, participants sitting in front of computers might not have as much true feelings of driving as using a driving simulator in the laboratory or in the real-world driving. The low-fidelity simulator experiment could lead to a false risk perception as compared with studies using driving simulators or real cars. Future works are recommended to examine the relationship between ToR lead time, driver SA and driver performance in the driving simulator and in the field - test to validate the web-based experiment.

The participants' response accuracy of SA queries was not very high on average, particularly for Level 2 and Level 3 SA, even when a long ToR lead time was given. It suggests some SA assistance technologies to be offered in the future design of conditional automation system, not merely extending the ToR lead time. For instance, augmented reality technologies could be equipped with the system to attract the driver's attention to environmental elements (Langlois and Soualmi, 2016); some operation-level guidance could be provided to the driver for understanding the situation quickly and taking a safe course of action after resuming manual control; a dynamic visualization of remaining driving distance or time before arriving at the freeway exit could be helpful for drivers making better takeover decisions before missing an exit.

5. Conclusion

The present study investigates the relationship between ToR lead time and drivers' situation awareness for resuming manual control from the conditionally automated vehicle to exit from freeways, as well as how the ToR lead time affects drivers' takeover reaction time and subjective evaluations. The findings showed that a ToR lead time of 16–18 s was most appropriate for exiting from a freeway after taking account of drivers' situation awareness and postexperience evaluations. Drivers used different strategies for regaining situation awareness and resuming control and therefore the takeover reaction time varied from person to person. For the drivers who prefer to take over control before acquiring adequate situation awareness, it could be detrimental when they are not ready for resuming manual control of the vehicle, and as a consequence, perform poorly after the takeover action. The future design of conditionally automated vehicle for freeway driving should consider approaches, such as equipping with SA assistance and assessment technologies, to assist drivers with obtaining adequate situation awareness before resuming manual control to exit a freeway.

CRedit authorship contribution statement

Xiaomei Tan: Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Yiqi Zhang:** Methodology, Software, Investigation, Formal analysis, Writing – original draft, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. (1850002) and PSU/Technion Marcus Funds. We would also like to thank Shujin Xu for helping with the coding of data in the analysis.

Appendix 1

1. About the car in front of you in your lane:
 - a. When the video stops, is the car in front of you in your lane turning its turn signal on?
 - ☐ Yes
 - ☐ No
 - ☐ I'm not sure
 - b. When the video stops, what is the car in front of you in your lane intending to do?
 - ☐ Pull over
 - ☐ Change lanes
 - ☐ Exit from freeway
 - ☐ Keep driving on freeway
 - ☐ I'm not sure
 - c. Now you have taken over control of your car (by pressing the spacebar), what do you need to do to cope with the car in front of you in your lane?
 - ☐ Follow the car and keep a safe distance
 - ☐ Change lanes to avoid possible collisions
 - ☐ No need to do anything
 - ☐ I'm not sure
2. About the car at your front right:
 - a. When the video stops, is the car at your front right turning its turn signal on?
 - ☐ Yes
 - ☐ No
 - ☐ I'm not sure
 - b. When the video stops, what is the car at your front right intending to do?
 - ☐ Pull over
 - ☐ Change lanes to exit
 - ☐ Keep driving on freeway
 - ☐ I'm not sure
 - c. Now you have taken over control of your car (by pressing the spacebar), what do you need to do to cope with the car at your front right?
 - ☐ Speed up and overtake
 - ☐ Slow down and let the car cut in
 - ☐ Change lanes to avoid possible collisions
 - ☐ No need to do anything
 - ☐ I'm not sure
3. About your car speed:
 - a. When the video stops, what is the speed of your car?
 - ☐ Less than 45 mph
 - ☐ 45-55 mph
 - ☐ 56-65 mph
 - ☐ 66-75 mph
 - ☐ More than 75 mph
 - ☐ I'm not sure
 - b. Is your car driving at an appropriate speed to exit from freeway (within a range of speed limit ± 5 mph)?
 - ☐ Yes
 - ☐ No
 - ☐ I'm not sure
 - c. Now you have taken over control of your car (by pressing the spacebar), do you need to brake to conform with the speed limit of the exit road?
 - ☐ Yes
 - ☐ No
 - ☐ I'm not sure
4. About your car's turn signal:
 - a. When the video stops, is your car turning the turn signal on?
 - ☐ Yes
 - ☐ No
 - ☐ I'm not sure
 - b. While driving in the current lane, do you need to signal your intention to exit from freeway?
 - ☐ Yes
 - ☐ No
 - ☐ I'm not sure
 - c. Now you have taken over control of your car (by pressing the spacebar), do you need to manually turn on the turn signal?
 - ☐ Yes
 - ☐ No
 - ☐ I'm not sure
5. The audio warning was sent XX mile away from the exit. The left end represents the location where you received the warning. The right end represents the freeway exit. Please drag the slider to the location where you pressed the spacebar.

Location where you receive warning		Exit (XX mile away)
------------------------------------	--	---------------------

6. How much were you ready to take over control of the car to exit from freeway?

Not ready at all				Neutral			Fully ready	
1	2	3	4	5	6	7	8	9

7. What do you think about the timing of the audio takeover request?

Extremely early (i.e., plenty of time for takeover)			Neither too early nor too late				Extremely late (i.e., too little time for takeover)		
1	2	3	4	5	6	7	8	9	

8. How much do you trust the takeover request system?

Not trust at all				Neutral			Extremely trust	
1	2	3	4	5	6	7	8	9

9. What was the workload for you in responding to the takeover request?

Extremely low				Neutral			Extremely high	
1	2	3	4	5	6	7	8	9

10. How much do you accept the takeover request system?

Not accept at all				Neutral			Extremely accept	
1	2	3	4	5	6	7	8	9

References

- Anwyl-Irvine, A.L., Massonnié, J., Flitton, A., Kirkham, N., Evershed, J.K., 2020. Gorilla in our midst: an online behavioral experiment builder. *Behav. Res. Methods* 52 (1), 388–407.
- Du, N., Kim, J., Zhou, F., Pulver, E., Tilbury, D.M., Robert, L.P., Yang, X.J., 2020. Evaluating effects of cognitive load, takeover request lead time, and traffic density on drivers' takeover performance in conditionally automated driving. In: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 66–73.
- Endsley, M.R., 1987. SAGAT: A Methodology for the Measurement of Situation Awareness (NOR DOC 87–83). Northrop Corporation, Hawthorne, CA.
- Endsley, M.R., 1988. Situation awareness global assessment technique (SAGAT). In: *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, pp. 789–795.
- Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. *Hum. Factors* 37 (1), 32–64.
- Endsley, M.R., 2019. A systematic review and meta-analysis of direct objective measures of situation awareness: a comparison of SAGAT and SPAM. *Hum. Factors* 0018720819875376.
- Eriksson, A., Stanton, N.A., 2017. Takeover reaction time in highly automated vehicles: noncritical transitions to and from manual control. *Hum. Factors* 59 (4), 689–705.
- Feldhütter, A., Gold, C., Schneider, S., Bengler, K., 2017. How the duration of automated driving influences take-over performance and gaze behavior. In: Schlick, C.M., Duckwitz, S., Flemisch, F., Frenz, M., Kuz, S., Mertens, A., Mütze-Niewöhner, S. (Eds.), *Advances in Ergonomic Design of Systems, Products and Processes*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 309–318. https://doi.org/10.1007/978-3-662-53305-5_22.
- Forster, Y., Naujoks, F., Neukum, A., Huestegge, L., 2017. Driver compliance to take-over requests with different auditory outputs in conditional automation. *Accid. Anal. Prev.* 109, 18–28.
- Gold, C., Körber, M., Lechner, D., Bengler, K., 2016. Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. *Hum. Factors* 58 (4), 642–652.
- Gugerty, L.J., 1997. Situation awareness during driving: explicit and implicit knowledge in dynamic spatial memory. *J. Exp. Psychol.: Appl.* 3 (1), 42.
- Heck, R.H., Thomas, S.L., Tabata, L.N., 2013. *Multilevel and Longitudinal Modeling with IBM SPSS*. Routledge.
- Heikoop, D.D., de Winter, J.C.F., van Arem, B., Stanton, N.A., 2018. Effects of mental demands on situation awareness during platooning: a driving simulator study. *Transp. Res. Part F: Traffic Psychol. Behav.* 58, 193–209.
- Holländer, K., Pfleging, B., 2018. Preparing drivers for planned control transitions in automated cars. In: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pp. 83–92.
- Kerschbaum, P., Lorenz, L., Bengler, K., 2015. June). A transforming steering wheel for highly automated cars. In: *2015 IEEE Intelligent Vehicles Symposium (iv)*, pp. 1287–1292.
- Jin, M., Lu, G., Chen, F., Shi, X., Tan, H., Zhai, J., 2021. Modeling takeover behavior in level 3 automated driving via a structural equation model: considering the mediating role of trust. *Accid. Anal. Prev.* 157, 106156.
- Körber, M., Gold, C., Lechner, D., Bengler, K., 2016. The influence of age on the take-over of vehicle control in highly automated driving. *Transp. Res. Part F: Traffic Psychol. Behav.* 39, 19–32.

- Langlois, S., Soualmi, B., 2016. Augmented reality versus classical HUD to take over from automated driving: an aid to smooth reactions and to anticipate maneuvers. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 1571–1578.
- Lorenz, L., Kerschbaum, P., Schumann, J., 2014. September). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? Proceedings of the Human Factors and Ergonomics Society Annual Meeting 58 (1), 1681–1685.
- Louw, T., Merat, N., Jamson, H. (2015). Engaging with Highly Automated Driving: To be or Not to be in the Loop? *8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Salt Lake City, Utah, USA*, 8.
- Lu, Z., Coster, X., de Winter, J., 2017. How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Appl. Ergon.* 60, 293–304.
- McCall, R., McGee, F., Mirmig, A., Meschtscherjakov, A., Louveton, N., Engel, T., Tscheligi, M., 2019. A taxonomy of autonomous vehicle handover situations. *Transp. Res. Part A: Policy Practice* 124, 507–522.
- Melcher, V., Rauh, S., Diederichs, F., Widroither, H., Bauer, W., 2015. Take-over requests for automated driving. *Procedia Manuf.* 3, 2867–2873.
- Metz, B., Wörle, J., Hanig, M., Schmitt, M., Lutz, A., 2020. Repeated usage of an L3 motorway chauffeur: change of evaluation and usage. *Information* 11 (2), 114.
- Miller, D., Sun, A., Ju, W. (2014). Situation awareness with different levels of automation. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 688–693.
- Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, automation off: unstructured transition timing for distracted drivers of automated vehicles. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems* (pp. 2458–2464). IEEE.
- Nobari, K.D., Albers, F., Bartsch, K., Bertram, T., 2020. Online feedback control for driver-vehicle interaction in automated driving. In: *International Conference on Applied Human Factors and Ergonomics*, pp. 159–165.
- Pampel, S.M., Large, D.R., Burnett, G., Matthias, R., Thompson, S., Skrypchuk, L., 2019. Getting the driver back into the loop: the quality of manual vehicle control following long and short non-critical transfer-of-control requests: TI: NS. *Theor. Issues Ergon. Sci.* 20 (3), 265–283.
- Petermeijer, S., Doubek, F., & de Winter, J. (2017). 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.
- Radlmayr, J., Gold, C., Lorenz, L., Farid, M., Bengler, K., 2014. How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proc. Human Factors Ergon. Soc. Annual Meeting* 58 (1), 2063–2067.
- SAE international. (2016). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE International, (J3016).
- Salmon, P.M., Stanton, N.A., Young, K.L., 2012. Situation awareness on the road: review, theoretical and methodological issues, and future directions. *Theor. Issues Ergon. Sci.* 13 (4), 472–492.
- Samuel, S., Borowsky, A., Zilberstein, S., Fisher, D.L., 2016. Minimum time to situation awareness in scenarios involving transfer of control from an automated driving suite. *Transp. Res. Rec.* 2602 (1), 115–120.
- Sukthakar, R., 1997. Situation awareness for tactical driving. Carnegie Mellon University. Doctoral dissertation.
- Thorn, E., Kimmel, S. C., Chaka, M., Hamilton, B. A. (2018). A framework for automated driving system testable cases and scenarios (No. DOT HS 812 623). United States. Department of Transportation. *National Highway Traffic Safety Administration*.
- van den Beukel, A.P., van der Voort, M.C., 2013. Retrieving human control after situations of automated driving: How to measure situation awareness. *Adv. Microsystems Automotive Appl.* 2013, 43–53.
- Van Den Beukel, A. P., & Van Der Voort, M. C. (2013b). The influence of time-criticality on situation awareness when retrieving human control after automated driving. *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2000–2005.
- Vlakveld, W., van Nes, N., de Bruin, J., Vissers, L., van der Kroft, M., 2018. Situation awareness increases when drivers have more time to take over the wheel in a Level 3 automated car: a simulator study. *Transp. Res. Part F: Traffic Psychol. Behav.* 58, 917–929.
- Wan, J., Wu, C., 2018. The effects of lead time of take-over request and nondriving tasks on taking-over control of automated vehicles. *IEEE Trans. Hum.-Mach. Syst.* 48 (6), 582–591.
- Wörle, J., Metz, B., Othersen, I., Baumann, M., 2020. Sleep in highly automated driving: takeover performance after waking up. *Accid. Anal. Prev.* 144, 105617. <https://doi.org/10.1016/j.aap.2020.105617>.
- Wright, T.J., Samuel, S., Borowsky, A., Zilberstein, S., Fisher, D.L., 2016. Experienced drivers are quicker to achieve situation awareness than inexperienced drivers in situations of transfer of control within a Level 3 autonomous environment. *Proc. Human Factors Ergon. Soc. Annual Meeting* 60 (1), 270–273.
- Yang, L., Semiromi, M. B., Auger, D., Dmitruk, A., Brighton, J., Zhao, Y. (2020). The implication of non-driving activities on situation awareness and take-over performance in level 3 automation. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society* (pp. 5075–5080). IEEE.
- Yun, Y., Oh, H., Myung, R., 2019. The effect of takeover lead time on driver workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63 (1), 1872–1873.
- Yun, H., Yang, J.H., 2020. Multimodal warning design for take-over request in conditionally automated driving. *Eur. Transp. Res. Rev.* 12, 1–11.