Exploring the benefits of conversing with a digital voice assistant during automated driving: A parametric duration model of takeover time

Kirti Mahajan[a], David R. Large [b], Gary Burnett [c], Nagendra R. Velaga [d*]

[a] Research Scholar, Transportation Systems Engineering, Indian Institute of Technology, Mumbai, India, kirtimahajan.nith@gmail.com

[b] Senior Research Fellow, Human Factors Research Group, University of Nottingham, Nottingham, UK, david.r.large@nottingham.ac.uk

[c] Professor, Human Factors Research Group, University of Nottingham, Nottingham, UK, gary.burnett@nottingham.ac.uk

[d] Associate Professor, Transportation Systems Engineering, Indian Institute of Technology, Mumbai, India, velaga@civil.iitb.ac.in (*Corresponding author)

ABSTRACT

Vehicle automation allows drivers to disengage from driving causing a potential decline in their alertness. One of the major challenges of highly automated vehicles is to ensure a timely (with respect to safety and situation awareness) takeover in such conditions. For this purpose, the current study investigated the role of an in-vehicle digital voice-assistant (VA) in conditionally automated vehicles, offering spoken discourse relating specifically to contextual factors, such as the traffic situation and road environment. The study involved twenty-four participants, each taking two drives (counterbalanced): with VA and without VA, in a driving simulator. Participants were required to takeover vehicle control following the issuance of a takeover request (TOR) near the end of each drive. A parametric duration model was adopted to find the key factors determining takeover time (TOT). Paired comparisons showed higher alertness and higher active workload (mean NASA-TLX rating) during automation when accompanied by the VA. Paired t-test comparison of gaze behavior prior to takeover showed significantly higher instances of checking traffic signal, roadside objects, and the roadway during the drive with VA, indicating higher situation awareness. The parametric model indicated that the VA increased the likelihood of making a timely takeover by 39%. There was also some evidence suggesting that male drivers are likely to resume control 1.21 times earlier than female drivers. The study findings highlight the benefits of adopting a digital voice assistant to keep the drivers alert and aware about the recent traffic environment in partially automated vehicles.

Keywords

Human-machine-interfaces, voice-user interfaces (VUI), conditional automation, SAE Level 3, passive fatigue, driver takeover.

1 INTRODUCTION

Recent advances in vehicle automation allow drivers to disengage from driving and become a passive monitor of the system, thereby bringing the potential risks of passive fatigue and associated road safety concerns to the forefront of current research (Matthews et al., 2018). The shift in driver responsibility from an active operator to a passive observer in an automated vehicle (AV) leads to the loss of active task engagement, thereby compromising drivers' alertness required to intervene in critical moments (Körber et al., 2015; Matthews et al., 2018; Neubauer et al., 2012; Sheridan and Parasuraman, 2002; Vogelpohl et al., 2018). Such a decline in alertness, caused by 'task-disengagement' or 'low workload' conditions during automated driving, induces a passive fatigue state, resulting in lack of task motivation and tiredeness (Matthews et al., 2018, 2012; May and Baldwin, 2009; Saxby et al., 2013).

As a consequence of passive fatigue, the drivers may lose awareness of the current traffic situation (Samuel et al., 2016). This is specifically a problem in partially autonomous vehicles (level 2 or level 3 automation according to SAE International, 2018), in which the system will issue a takeover request (TOR) to the driver in situations that fall outside of its capability. In addition, drivers of level 4 automated vehicles may actively choose to take over control at some point (SAE International, 2018). An additional concern is that drivers in level 3 and level 4 automated cars are able to engage in secondary non-driving related tasks (NDRT) (SAE International 2018), and these may further disengage or fatigue the driver. Ensuring a safe and timely takeover of vehicle control is therefore one of the major challenges for highly automated driving. Consequently, it is suggested that some form of system feedback is required during periods of automation, to maintain driver alertness with appropriate situation awareness, prior to taking over vehicle controls (Wu et al., 2019).

1.1 Automation and alertness

Automation in vehicles can lead to prolonged periods of driver inactivity resulting in significant loss of alertness, with some studies recording a significant increase in symptoms of fatigue after only 15-20 minutes of automated driving in an AV (Gonçalves et al., 2016; Neubauer et al., 2012; Vogelpohl et al., 2018; Wu et al., 2019). Gonçalves *et al.* (2016) noticed that after only 15 minutes of automated driving there was significant increase in the degree of sleepiness among the majority of drivers, indicated using the Stanford Sleepiness Scale (SSS; Hoddes et al., 1973). Neubaue*r et al*. (2012) found significant correlations between the fatigue ratings and lower task engagement during automation– indicated by the Dundee Stress State Questionnaire (DSSQ; Matthews et al. 2002). Saxby et al. (2013) also observed that automation led to lower workload ratings on NASA-TLX scale (Hart and Staveland, 1988; Matthews et al., 2013) resulting in passive fatigue. Wu et al. (2019) reported a significant increase in eye blink durations and Karolinska Sleepiness Scale ratings, (KSS; Åkerstedt and Gillberg, 1990) indicating fatiguing effects after 30-minutes of automated driving, and these effects were irrespective of the driver's age. Vogelpohl *et al.* (2018) noticed same level of fatigue among drivers after 20 minutes of monotonous automated driving, as compared to 40-50 minutes of manual driving. Collectively, these studies suggest that a period of automated driving lasting 20 minutes or longer is sufficient to significantly lower task engagement and cause a significant decline in driver alertness.

1.2 Takeover time

Partially automated vehicles have a specific domain of automated operations, beyond which they require human intervention. This is conveyed through pre-defined alerts, referred to as a takeover request (TOR) made to the driver. The takeover time (TOT) is the response time of drivers to the TOR. It includes both the time it takes a driver to make an assessment of the traffic situation, i.e. regain their *situation awareness (SA)* (Vlakveld *et al.*, 2018), and demonstrate their readiness to drive by re-engaging with the driving controls (Gold et al., 2018, 2013). Many studies have used the ability to respond to emergency situations or hazard perception as a good proxy measure of situation awareness during the takeover process (Gold et al., 2013; Samuel et al., 2016; Zeeb et al., 2015). In this context, the general approach is to determine the takeover time and its quality by measuring the driver's behavior in response to a hazard scenario occurring immediately after takeover, for example, by decelerating or changing lanes to avoid a braking vehicle ahead (Gold et al., 2013; Neubauer et al., 2014; Samuel et al., 2016; Wu et al., 2019; Zeeb et al., 2015). The majority of these simulator studies found that drivers "comfortably" completed the process of taking over controls post TOR in less than 10s (Gold et al., 2013; Gonçalves et al., 2016; Mcdonald and Alambeigi, 2019; Vlakveld et al., 2018; Vogelpohl et al., 2018; Wan and Wu, 2018). However, in the case of a potential collision event, the response is often reflexive, typified by sudden, emergency braking (Gold et al., 2018; Zeeb et al., 2015). This does not necessarily indicate that drivers are fully prepared to control their vehicle and engage with other road users. To more accurately measure the time to resume manual driving when a TOR is issued, it would therefore seem prudent to avoid situations inviting a reflexive response, and instead select a more authentic or routine driving scenario. This approach was taken by Merat et al. (2014), who reported that an average time to respond to resume steering and brakes in response to TOR was between 10 and 15 seconds, where the drivers responded at their ease in the absence of any hazard detection event at TOR. A brief review of relevant studies is presented in Table 1.

1.2.1 Factors influencing takeover time

Several studies have found different factors influencing takeover time after highly automated driving (Gold et al., 2013; Louw et al., 2017; Merat et al., 2014; Zeeb et al., 2016, 2015). In general, the response time of drivers is influenced by various factors. For instance, older drivers are known to respond slower than younger drivers (e.g. Scullin and Bliwise 2015). However, a quantitative review by Zhang et al. (2019) revealed no significant effect of age on takeover time and concluded that the probability of compensatory effects of driving experience alleviated cognitive decline due to age (Shinar et al., 2001; Smith et al., 2009). In general, previous studies indicate a difference in reaction time between males and females (Lings, 1991; Mehmood and Easa, 2009; Warshawsky-livne and Shinar, 2002). Gomez et al. (2019) found that females tended to exhibit more exploratory visual behavior, including longer fixations and scan path lengths during a task, and this could potentially lengthen the takeover time. Nevertheless, Zeeb et al. (2015) studied the process of takeover using driver's visual behavior and found that age, gender, mileage and experience with driver assistance systems did not influence visual behavior. Even so, these demographic factors are often reported to influence the response time of fatigued drivers (Dozza, 2013; Mahajan and Velaga, 2020; Smith et al., 2009), highlighting the need to consider these in the context of taking-over control of an automated vehicle.

1.2.2 *Visual behavior as an indicator of situation awareness*

Automation causes mental underload situations which can potentially result in passive fatigue and signs of sleepiness, and is therefore likely to reduce the visual attention of drivers (Vogelpohl et al., 2018; Zeeb et al., 2015). It therefore follows that responding to a takeover request in an AV may require additional time as drivers are required to regain self-alertness before building their situation awareness (SA) and responding to the TOR (Vogelpohl et al., 2018; Zeeb et al., 2015). Visual, or gaze behavior can be a useful method to explore the process of allocating attention (Young and Stanton, 2002) (also, see Table 1). Factors such as the duration and frequency of glances spent checking the speedometer, the road ahead, and side and rearview mirrors, are associated with building situation awareness (Gold et al., 2013; Samuel et al., 2016; Vlakveld et al., 2018; Wright et al. 2016). Furthermore, actions such as placing hands on steering, placing feet on the pedals and looking at road ahead are indicative of motor readiness or readiness to drive (Gold et al., 2018, 2013; Zeeb et al., 2015). Gold et al. (2013), Neubauer, Matthews, & Saxby (2014) and *Merat et al.* (2014) reported driver's involvement in checking rear and side mirrors to gain situational awareness in response to a TOR (the TOR was followed by a collision event in these studies). Vogelpohl *et al.* (2018) reported that the first glance to speedometer (contained in the dashboard) was reported at 14 to 15 seconds after the TOR under passive fatigue conditions. Gonçalves, Happee and Bengler (2016) provided a TOR with a limiting time budget of 5 seconds (triggered by a SSS sleepiness rating of 5 or more), although eye-based sleep indicators or frequency of checking mirrors did not provide any evidence of drowsiness due to automation. Collectively, this highlights the need to develop systems to keep the drivers engaged and updated with driving and the traffic environment specifically during periods of automation, thereby improving the takeover response.

1.2.3 *Modelling takeover time*

The literature suggests various covariates which affect the takeover time and performance (safe or unsafe) (also, see Table 1). To determine the contribution of different factors (driver demographics and specifically, gaze behavior) and their impact on the time to motor readiness in a takeover situation, it would be beneficial to model the takeover time. Zeeb et al. (2015) adopted an integrated model approach to emphasize that primarily cognitive processes such as gaze behaviour determine the takeover time. However, the study findings were limited to an emergency scenario presented at takeover and did not accommodate for the effect of other covariates such as gender, driving experience etc. as discussed previously. Existing studies in transportation research have widely used a parametric duration modelling approach to model the response time of drivers in emergency situations (Choudhary and Velaga, 2017a; Haque and Washington, 2014; Mahajan and Velaga, 2020; Yadav and Velaga, 2019). This is used to determine the conditional probability of the elapsed time until the duration of an event of interest, provided the event does not end in this duration (Washington et al., 2003). The advantage of using such a modelling approach is that it accounts for the effect of various covariates on the response time.

Table 1 Literature review of takeover time and passive fatigue and related measures.

Citation	**Variables studied**	**Sample size**	**Instrument used**	**Measures**	**Analysis technique**	**Findings**
(Gold et al., 2013)	Warning time for TOR (5s versus 7s) Hazard scenario	37 (AV) and 13 (manual)	Gaze detection in TOR scenario	• RT: Acceleration, brakes, Steering • Checking mirrors (glances)	t-test comparison	With shorter TOR-time, the subjects react faster with excessive use of brakes and reduced mirror checks etc.
(Wu et al., 2019)	Emergency event after 6s of TOR to test RT (Verbal TOR)	115	EOG KSS Eye blink duration	• Minimum time to control either pedals or steering	t-test ANCOVA	The signs of fatigue reported after 4-6 min of automated drive and 15 minutes of automated driving indicated by KSS and avg. eyeblink duration. Emergency event raised alertness.
(Neubauer et al., 2012)	• Hazard scenario • Optional automation	93(AV) and 91 (NA)	DSI, DSSQ and NASA-task load index	• SDLP (at 30s for 2 min) • RT to avoid collision: brakes, Steering	t-test ANOVA	• Based on analysis of DSSQ, drivers experienced fatigue and stress after 10 minutes of drive. • Optional automation did not improve driver's alertness or performance.
(Merat et al., 2014)	Fixed transfer versus Variable transfer	37	Eye tracking and head movement	• SDLP • Frequency of SRR $\geq 1^{o}$ • Gaze: percentage road center PRC:	ANOVA	Values in 1st min of manual control, SDLP and gaze focus showed a lag of 10-15s while steering control lagged 35-40s in resuming control after automation stopped. So the TOR shall be >40s prior to control transfer.
(Gonçalves et al., 2016)	Between-subjects study: Reference group (G1) Vs drowsy group, G2 (Drowsiness (SSS>5) triggered TOR)	16 (G1) +15 (G2)	Eye tracking Facelab 5.1	• SWA and acceleration • Mirror checking	ANOVA	Mean 15 minutes of automated drive can cause drowsiness among drivers.
(Vogelpohl et al., 2018)	• Sleep deprived and time of day • Long drive and monotony	60 (15 in 4 groups)	Facial, eyebased indicators reported by observer	• KSS post drive • RT to disengage automation, hazard (braking, steering or overtaking) • Situation awareness:	ANOVA	• AV caused medium to high level fatigue among 50% drivers after 20 min while it took 40-50min in non-AV to reach comparable fatigue level in same ratio of drivers. • Fatigue increased the RT (5-8s) to TOR in AVs
(Vlakv*eld et al.*, 2018)	Comparison of TOT (4s vs 6s)	21(4s) + 22(6s)	Eye-tracking	• Glance frequency at latent hazards (SAR) • Time to disengage-auto	ANOVA	Group with longer TOT (6s) gazed at more number of latent hazards than 4s group.
(Samuel et al., 2016)	Latent hazard scenario at TOR		Eye tracking	• RT	logistic regression	Longer TOT (8s and 12s) improved latent hazard perception of drivers comparable to manual driving condition.
(Zeeb et al., 2015)	Hazard scenario (Verbal message and Visual icon)	N= 89 (54M, 35 F)	Eye-tracking	• Frequency and duration of glances • AOI analysis • RT to hazard	k-means clustering ANOVA	• Accidents were more among drivers with more proportion of time spent off-road. • Mileage, experience, age or use of adaptive cruise control and lane keeping assist did not affect gaze behavior at TOR.

1.3 Conversation to counter passive fatigue

Various studies have claimed the alerting effects of engaging in an active conversation for the drivers while driving (Large et al., 2018; Neubauer et al., 2014; Young, 2013). However, very few studies have been conducted to extend such benefits (i.e. of engaging in conversation) to mitigate the effects of passive fatigue during automated driving, and in particular, the impact on the time to gain motor readiness. For instance, Saxby, Matthews, & Neubauer (2017) engaged drivers in short 30 second telephone conversations with the experimenter to share their 'experience of a traffic collision' with the aim of keeping drivers alert. However, such content evidently distressed the drivers and negatively influenced their post takeover driving performance. Thus, they suggested that a longer conversation on more suitable topics might prove more effective in maintaining driver alertness during automated driving. In addition, it has been found that intermittent conversation tends to increase drivers' workload and alertness to an appropriate level during monotonous automated driving conditions, compared to continuous talking which may distract the drivers (Atchley et al., 2013). The majority of these studies used cell phone conversations. However, there is an inherent risk of distraction and performance deficits caused by cell phone conversations (Choudhary and Velaga, 2017b, 2017c, 2017a). Alternatively, a few studies have proposed the use of a digital voice assistant (VA) or a natural language interface conversing with the driver throughout the journey (Antrobus et al., 2018; Large et al., 2018, 2017, 2016). These studies found that short intermittent conversations with a digital VA improved driver alertness and avoided potential driving performance decrements due to low alertness (Large et al., 2018). Large et al. (2016) showed that normal conversations about calendar reminders, news, radio or music etc. with the VA were less cognitively demanding than a cell phone conversation but were equally effective in maintaining the alertness of drivers during fatigued manual driving conditions. Drews, Pasupathi and Strayer, (2008) found that a naturalistic conversation between the driver and their passengers about the surrounding traffic mitigated the distracting effect of conversation. In such a case, employing a natural language interface – conceptualised as a voice assistant, could replace the passenger, and provide timely feedback to the driver about recent or upcoming traffic situations during automated driving. Such information may help the driver to effectively maintain driver alertness throughout the journey and may additionally reduce the takeover time as drivers will already be aware of relevant aspects of the driving situation.

1.4 Study motivation and hypothesis

Some studies have highlighted the need of a system in future vehicles that can provide intermittent feedback about the traffic situation around the drivers to help them resume control (Drews et al., 2008; Merat et al., 2014; Vlakveld et al., 2018). Using this approach, this study aimed to examine how a digital voice assistant can help in mitigating passive fatigue induced by automation and in improving situation awareness at takeover. In addition to providing traffic-related information, the nature and content of the conversation were also inspired by previous studies using a natural language interface (Antrobus et al., 2018; Large et al., 2018, 2017, 2016). In this study, however, it is hypothesized that intermittent, traffic-related conversation with a voice assistant (VA) will reduce delays in takeover time caused by passive fatigue and disengagement from driving during highly automated driving. Nevertheless, providing traffic narrative at the point of takeover is likely to distract the driver and could influence their driving performance and should therefore be avoided. According to the previous studies already discussed, gaze behavior, such as frequency or duration of checking mirrors and the road ahead, are indicators of building/updating awareness of the surrounding traffic and driving situation. Therefore, we hypothesize that the traffic status updates provided by VA prior to the TOR, will help redirect drivers' attention to the road, traffic or traffic signages, as per the messages, which can be confirmed through their gaze behavior. Thirdly, there are other factors influencing the takeover process, including driver demographics. The previous studies had investigated the influence of each of these factors (age, experience, gender, involvement in secondary tasks, etc.) in isolation. However, in reality, such factors act simultaneously to influence the takeover process. Further, the effectiveness of a voice-based system – such as that proposed here, also depends on the ease and interest of drivers to use such technology. Therefore, this study focuses on modelling the takeover time to quantify the relative contribution of all such covariates, thereby evaluating the effectiveness of a VA in assisting takeover after automated driving.

1.4.1 The present study

Based on these, the study aims to address the following questions:

1) Can engaging in intermittent conversations with a digital voice assistant avoid (or mitigate) the loss of alertness during automation?
2) Can verbal traffic cues redirect drivers' visual attention and help in building their situational awareness prior to takeover?
3) How does the presence of VA (directly or indirectly in terms of cognitive processes such as gaze behavior, driver age, gender or accustomed to use of other assistants), influence the time to gain motor readiness when a TOR is prompted by the system?

The methodology adopted to achieve the desired aims in this study is discussed in the following section.

2 METHOD

The study was conducted using a high-fidelity, fixed-base driving simulator at the University of Nottingham Human Factors laboratory (Figure 1a). This driving simulator comprises an Audi TT car located within 270 degrees field of view provided by a curved screen and overhead HD projectors. Three inobtrusive cameras were installed at different positions inside the car to capture drivers' hand and feet movements as well as any physical signs of sleepiness during the experiments (yawning, extended blinks etc.). This simulator has been used in various previous relevant studies (Antrobus et al., 2018; Large et al., 2018, 2017, 2016), and is capable of providing an experience of driving level 3 automated vehicle.

2.1 Participants

Participants were invited to take part using flyers displayed around the University campus. They were required to be active and experienced drivers, without requiring any corrective eye-glasses to drive (to ease collection of eye-tracking data), and shall not suffer from any sleep-related disorder or excessive daytime sleepiness. Participants' suitability was evaluated using a pre-study questionnaire. In line with similar previous studies, excessive daytime sleepiness was determined using the Epworth Sleepiness Scale (ESS; M.W. Johns 1990) i.e. ESS score ≥16 (Philip *et al.*, 2002; Miletin and Hanly, 2003; Lomelí *et al.*, 2008). None of the volunteers reported any sleep-related disorder or excessive daytime sleepiness. However, one volunteer was excluded for not having a valid driving license and two volunteers were excluded as they wore correcting glasses. In addition, participants were instructed to refrain from consuming caffeine, mints or alcohol for a few hours immediately prior to the study, and to take adequate sleep the day before the study – participants were also reminded of the latter requirements 24 hours before they were due to attend, in line with previous studies on driver fatigue (Centofanti et al., 2017; Hilditch et al., 2017; Otmani et al., 2005; Smith et al., 2009; Zhang et al., 2016).

Thirty-one eligible participants volunteered for the study. Unfortunately, three participants failed to arrive for the experiment and four stopped partway through due to suspected simulator sickness. The data from the remaining twenty-four participants (10 females and 14 males) is included in the analysis. Participants were aged between 22-60 years (mean = 30.13, SD = 8.39) and driving experience ranged from 2 to 33 years (mean= 10.5, SD = 6.89). All the drivers were thanked for their participation with a £20 gift voucher. The study protocol was approved by the Faculty of Engineering Research Ethics Committee, University of Nottingham, and all participants provided written informed consent before taking part. To encourage more natural behaviour during automation, participants were not pre-informed about the actual/precise aim, i.e., to study 'the passive fatigue due to automation'. As per the approved study proposal, the actual aim of the study was only revealed at the end of the study. Therefore, each participant also signed a written consent form at the end of the study to confirm their understanding and receipt of compensation voucher.

2.2 Experimental Set-up

The study involved a within-subject design with two driving sessions – one with and one without the voice assistant. A bespoke scenario was created using STISIM Drive 3 software. The route represented a transition from an initial two miles of two-lane urban road to a standard UK dual carriageway (Figure 1b) and back to a similar urban scene. The speed limit was posted between 30-40mph in the urban scene and 50-70mph in the dual carriageway, in line with normal UK road conventions. The same driving scenario was used in both the sessions with minor changes in the environment such as different buildings, types of cars etc. The roadside environment was intentionally sparse, with limited additional traffic, to minimise auditory and visual stimuli and increase monotony. Each drive lasted approximately 30 minutes in the driving simulator with approximately 25 minutes of automated drive – considered to be sufficient to induce passive fatigue based on similar research settings (Wu et al., 2019; Neubauer *et al*., 2012; Gonçalves, Happee and Bengler, 2016 and Vogelpohl *et al.*, 2018).

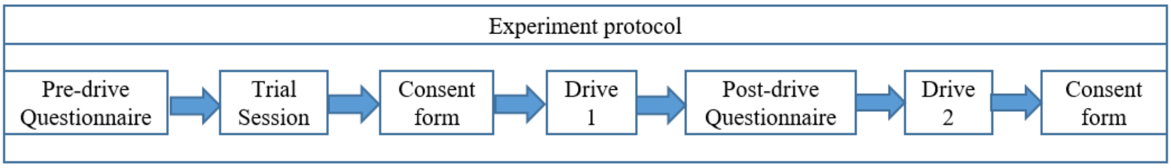
a.
b.

Figure 1 a. Driving Simulator set up and b. general design view of dual carriageway

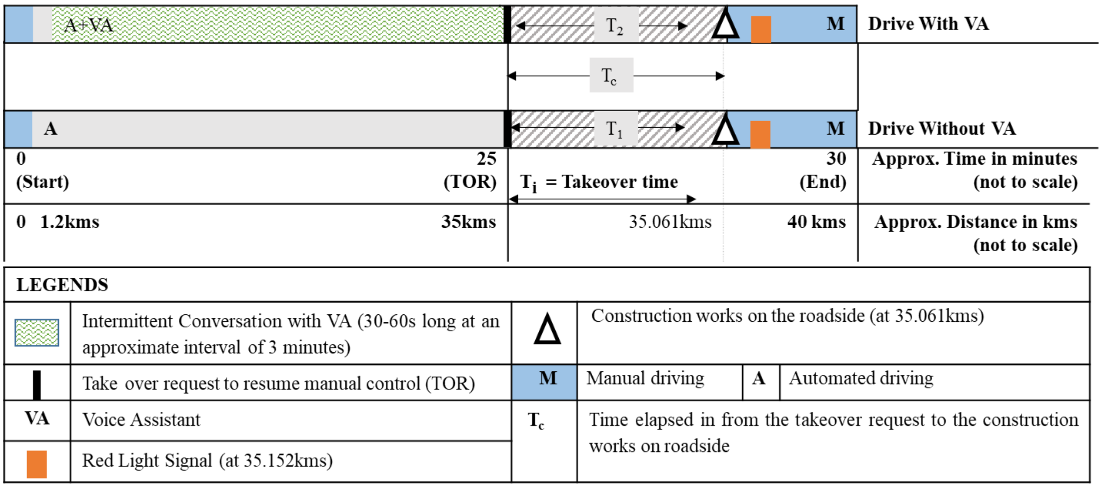
All participants drove in both the driving conditions (with and without VA), which were counterbalanced to avoid order effects on takeover performance (Sarkies et al., 2019). In addition, the surroundings including buildings, types of cars etc. were changed in the two drives particularly when the drivers entered the two-lane urban road again prior to takeover event to avoid prediction of the TOR at any location based on identifiable surroundings or landmarks. A general layout of the study protocol and experiment design in each drive is illustrated in Figure 2.

2.2.1 Drive with Voice Assistant (VA)

During one of the two drives (which were counterbalanced), drivers were provided with the VA, which instigated and enabled natural language interactions. The aim was to keep the driver engaged during the drive and thereby avoid fatigue. VA was designed to provide useful information to the drivers, such as surrounding traffic feedback, route navigation, event reminders or offering to play some music or radio during the drive (refer Table 2). The VA introduced itself and its capabilities to the driver prior to the start of the drive. Drivers could either respond to or initiate discussion with VA using natural, conversational language. To achieve this, a comprehensive set of pre-recorded spoken messages were embedded in the STISIM scenario, and these were played as required by the experimenter during the experiment. Drivers could also initiate a conversation with VA. In situations where drivers initiated conversation concerning topics for which no specific responses had been recorded, for example, requesting a piece of music or information about a particular topic, VA responded with an error message, such as "sorry! no network connectivity at the moment to perform this task", "sorry! This function is not available currently". In practice, these messages were rarely used.



a. Experiment protocol for the study involving two drives with each participant



b. Representation of a general layout of each drive (with or without VA) conducted on the driving simulator

Figure 2 A schematic representation of the study protocol including two drives

The first conversation topic was played after 5-minutes from the start of the drive, based on the expected onset of fatigue symptoms after 5 to 7 minutes of automation (Wu et al., 2019). Each participant received the same opening gambits. However, the follow up statements differed slightly based on each individual's response. VA began a new conversational exchange or topic at intervals of no more than 3 minutes to avoid prolonged periods absent of speech and to keep the drivers alert i.e. either a question or new information was conveyed to the drivers at a maximum interval of 180s. However, since the conversations were inspired from traffic and environment and the drivers could also initiate the conversation to their willingness, therefore, there could be a few conversations at intervals shorter than 3 minutes. The duration of most of these conversational exchanges lied between 30 and 60 seconds, as recommended by Saxby et al. (2017).

The conversation was primarily designed to provide traffic and road-environment related information (examples presented in Table 2), such as upcoming changes to speed-limits, but also included general conversation, as used in previous, similar studies (e.g. Large et al., 2018). The general conversations (refer to category 1 and 2 in Table 2) were not included at any fixed locations. For example, the entertainment related conversation was sometimes initiated by the drivers and VA responded. However, the traffic/road environment related conversations were always triggered by the experimenter based on nearby traffic signages or information sign boards. For example changing speed limits near (i.e. either just before or after the posted sign) a speed limit board, route information (refer Column 3b, Table 2) was played near a route information board or sometimes in response to driver's query 'where are we currently?', fatigue related conversation (refer column 3a, Table 2) were triggered by the experimenter near to a variable message sign (VMS) displaying the warning: ‘Tiredness Kills’, and the follow up conversation regarding facilities at service station was conveyed prior to the ‘service station’ sign board. In a random section of the roadway with no/low traffic, a vehicle in the adjacent lane slowly shifted to the driver’s lane, to which the automated vehicle maintained a safe gap. However, the VA asked to confirm that the driver was comfortable with the same choice of speed (refer conversation 3c, Table 2). These conversational exchanges were intentionally placed near signages, so that the driver was able to confirm the information conveyed by VA – with the aim of helping the driver build trust in the device. Also, the driver’s response to any question would determine the extent of follow up questions and length of conversation. For example, if the driver responded ‘No’ to playing music, the conversation would shift to offer to play a talk-based radio station instead and may subsequently end sooner. As the conversations were pre-recorded messages, each conversation was therefore limited to last between 30 to 60 seconds depending on the driver’s response. There was no conversation initiated during the final minute of automation, i.e. 60 seconds prior to the TOR, although VA had already informed drivers about the upcoming change in the posted speed-limit and the approaching pedestrian crosswalk.

Table 2 Examples of VA’s opening gambits

Category	Statements
1. Event reminders (calendar)	*“I have set a reminder you are currently ahead of schedule. I've also looked at your to-do list and messages. There are currently no messages left for you”*
2. Entertainment	*“Would you like me to play some music or radio for you? …Would you like me to play anything in particular? ... I can save your preferences; would you like to save it in your favourites? ...Okay! Saving in favourites. Playing music. [Music playing]”*
3. Traffic/ Road	
(a) Fatigue/Rest	*“You have been driving for a long time today, Would you like to stop for refreshments or rest? …There is a service station at the next right exit approaching in 10 minutes. …There are refreshments and lodging facilities also nearby, would you like me to make any reservations?”*
(b) General journey	*“We are currently headed towards ‘A5250’. The posted speed limit is ‘60’ miles per hour. Weather expected is clear and sunny in next 2 hours.”*
(c) Road-traffic feedback	*“You are too close to the leading car. Are you comfortable with the speed or Would you like to slow down?”… “Reducing speed. You are now driving at ‘55’ miles per hour.”… “There is a pedestrian crosswalk ahead. Please be engaged in the drive or would you like to slow down?”*

2.2.2 Takeover event

Prior to the test drives, participants undertook a practice drive so that they were familiar with the transfer of control from ‘manual’ to ‘automation’, and vice-versa. The instruction to transfer control was conveyed with a voice message followed by three consecutive beeps indicating the precise moment of the transfer of control, thereby avoiding any visual distraction. Drivers were presented with multiple instances of takeovers during the practice drive to gain some experience of switching controls in the automated vehicle. During the test drives, participants were pre-informed that they may be required to resume manual control at a certain point, but otherwise, should relax (Wan and Wu, 2018). After completing 1-minute into the drive, automation was engaged at a fixed point at 0.75 miles (1.2km) in each test scenario. Towards the end of the drive – after approximately 21.6 miles (35km) – participants received a takeover request (TOR). In the real world, a partially automated vehicle is likely to present a takeover request in sudden or unanticipated situations (though not necessarily a "hazard" situation). Therefore, to emphasize the need of a timely response to the prompted TOR without imposing a critical hazard, a construction zone was created on the roadside (without significantly intervening with the traffic flow i.e. no lane change was required). This occurred at about 61m (200ft) from the onset of TOR and comprised a gravel pile spilling near the left road edge (Figure 3). A distance of 61m (200ft) was selected through various pre-trials to allow the drivers ample time to respond to TOR. Also, it was sufficiently larger than the required braking distance under normal conditions (thereby avoiding the need for emergency braking) when approaching with a speed of 40mph. The construction zone was followed by two consecutive traffic signs displaying changing speed limit at approximately 30m (100ft), a pedestrian crossing sign at 61m (200ft), and a red-light traffic signal with a pedestrian cross-walk was presented at about 92m (300ft) from the construction zone, i.e. 152 m (approximately 500ft) from the point of TOR. The drivers were naturally expected to resume control and apply their brakes in response to the signal. For a device like VA to present traffic feedback and information, it must be a part of an integrated, intelligent transportation system, which would also include maps and navigation abilities. Therefore, during the drive with VA, drivers received a prior notification a few minutes before the TOR, that they were approaching a pedestrian cross-walk, whereas the construction zone at the roadside was presented as a sudden/unanticipated event to present a more realistic need to takeover control.

With the intention to measure driver situation awareness, visual allocation of driver attention to various components in one scene were analysed near the takeover. Therefore, the traffic signs, signal, construction zone (roadside objects) etc. were created as components of the driving scene at the takeover event. The required data to understand the takeover process (situation awareness and alertness) were collected at the takeover event. Manual driving behaviour and performance following the traffic signal was not analysed as it was assumed that drivers had regained alertness by the fact that they had successfully stopped at the traffic signal.

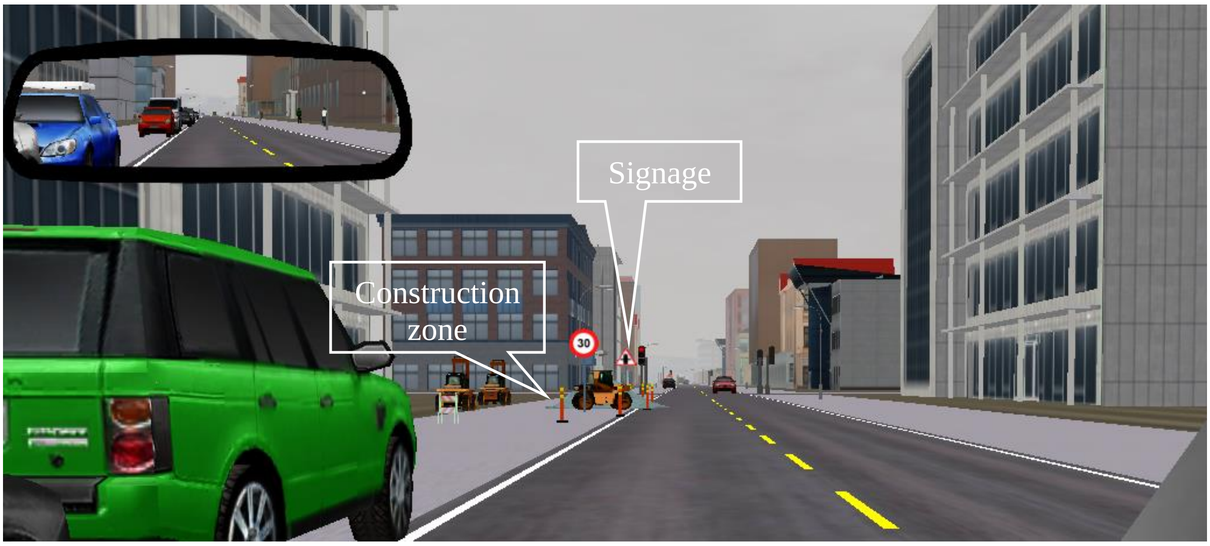


Figure 3 Scenario at takeover request (design view in STISIM at 100ft or 30.5m after onset of TOR)

2.3 Measures

A range of subjective and objective measures were obtained to assess the changes in driver alertness due to automation and their consequent performance in response to the TOR in each drive.

2.3.1 Visual characteristics

SMI 'natural gaze' eye-tracking glasses were used to collect visual behavioral data at a sampling rate of 30 Hz. The visual characteristics or eye-tracking data was collected as: 1) objective measures of alertness such as pupil diameter, eye blink frequency and eye-blink durations and 2) glance behavior as a measure of situation awareness, as employed in other similar studies (Gold et al., 2013; Gonçalves et al., 2016; Jamson et al., 2013; Körber et al., 2015; Large et al., 2018; Lu et al., 2017; Merat et al., 2014; Vlakveld et al., 2018). An increase in the frequency of eye blinks or pupil diameter are common indicators of fatigue or drowsiness while driving (Gonçalves et al., 2016; Large et al., 2018; Merat et al., 2014; Wu et al., 2019). Thus, the influence of VA on driver alertness during automation could be determined by comparing these measures for the two drives.

In order to determine drivers' awareness of the road environment (i.e. their 'situation awareness'), immediately prior to the TOR, glance duration and frequency were calculated for defined areas of interest (AOIs) using semantic gaze mapping (with BeGaze 3.7 software); a glance is defined as a fixation lasting for 200ms or over (Kapitaniak et al., 2015). The following AOIs were selected which were expected to be influential in determining the situation awareness of drivers: external mirrors (side-view mirrors), rear mirror, road ahead of the driver (windshield), roadside objects (obstacles), speedometer, traffic signs and signal in line with similar studies (Gold et al., 2013; Körber et al., 2015; Lu et al., 2017; Merat et al., 2014; Vlakveld et al., 2018). The remaining glances at areas other than the defined AOIs, such as glancing at car-interiors, inside car cameras, etc. were categorised as "others". Visual behavior was analysed during the 60 seconds prior to the red-light stop signal, which appeared approximately 195m (640ft) following the onset of TOR voice message (similar to a study by Zeeb et al. 2015).

Previous studies have used both count and duration of glances to study the allocation of visual attention on various components of driving scene or their contribution to situation awareness, such as, traffic signs (Babi et al., 2020; Campagne et al., 2005; Martens and Fox, 2007), external or rear view mirrors (Gold et al., 2018; Lu et al., 2017; Zeeb et al., 2015), in-vehicle systems (Kohl et al., 2020; Starkey et al., 2020), monitoring the road ahead (NHTSA, 2006; Noble et al., 2021) and other areas of interest. Larger glance duration indicates higher processing time to understand any given stimuli (Martens and Fox, 2007) such as visual messages, information signs etc. or may even indicate distraction (longer glances (>2s) away from the forward roadway) (NHTSA, 2006; Simons-Morton et al., 2014). Glance frequency on the other hand suggests frequency of scanning the road environment or its components (NHTSA, 2006). Unfamiliar objects (for example the construction zone at the roadside in this study) or unfamiliar traffic signs may need longer processing time which can be achieved either through frequent glances of short duration or through few glances of longer fixations (Babi et al., 2020; Campagne et al., 2005; Martens and Fox, 2007). However, less frequent yet longer fixations away from the road may also disengage the drivers from the primary task of monitoring the roadway (Kohl et al., 2020). Therefore, both the proportion of glance duration and the frequency of glances at various AOIs were recorded and analysed in this study to determine the differences in glance behavior due to VA.

2.3.2 Subjective sleepiness and workload ratings

As indicated in Table 1, previous studies have used different indicators of driver fatigue such as self-reported subjective scales, physiological or eye-tracking metrics in combination with each other to indicate driver fatigue or changes in alertness (Gonçalves et al., 2016; Merat et al., 2014; Vogelpohl et al., 2018; Wu et al., 2019). Therefore, in addition to the visual indicators of fatigue, drivers rated their level of alertness using the Karolinska Sleepiness Scale (KSS) and cognitive workload using NASA-TLX workload index, (Large et al., 2018; Matthews et al., Neubauer et al., 2018; Neubauer et al., 2012). KSS is a subjective scale which records the self-perceived value of alertness on scale of 1 (extremely alert) to 9 (extremely sleepy, sleep onset soon). The subjective ratings on KSS are closely associated with physiological indicators of fatigue i.e. EEG and EOG activity /drowsiness (see Lal and Craig, 2001 for a review). Wu et al., 2019 showed that KSS ratings and eye blink duration were highly correlated to each other and indicated increasing fatigue due to automation. NASA-TLX allows the drivers to rate the workload based on six factors i.e. mental, physical, temporal, performance, effort and frustration experienced by the them during a task on a increasing scale of 1 to 21(Hart and Staveland, 1988).

The subjective KSS ratings were collected on three occasions during each drive: firstly, prior to each drive, secondly, towards the end of automated drive (prior to TOR) and finally, after resuming manual drive. To avoid any interference by the experimenter during the drive, the latter two ratings were both collected at the end of each drive. Both of these ratings were collected based on drivers' memory of alertness felt by them just before takeover and at the end of experiment after 5 minutes into the manual driving post takeover. As both ratings were collected at the end of the drive, it is possible that some drivers failed to accurately recall their level of alertness immediately prior to the takeover. However, the intention in adopting this approach was to avoid interrupting the drive as this would have had the unintended consequence of increasing alertness. It is also worth noting that the takeover was fixed in that it did not require the drivers to reach any pre-decided level of change in alertness to trigger the TOR and therefore, no such rating was collected partway through automation. The three ratings were then compared within each drive and across the two drives – with and without the VA. The experimenter also noted the relevant symptoms of sleepiness e.g. frequency of yawning and incidents of micro-sleeps (or 'nodding off') behind the wheel, when automation was engaged.

2.3.3 Time to takeover request (TOR)

The response time to the TOR was measured as the time to resume driving controls. In practice, this was calculated as the time from the start of the stimulus (i.e. end of TOR voice message and start of beeps) to the time at which drivers acquired motor readiness i.e. hands on steering, feet on pedals and looking ahead (or eyes on the road). The time to resume steering and pedals were determined using frame-by-frame analysis of videos captured during the drive at a frame rate of 30 fps, whereas the time to resume glances on road was noted from the eye tracking videos played at the same frequency of 30Hz. Both the videos were also matched with each other for confirmation by displaying the frame number on each video. The response times (RT) were calculated as the difference in frame number corresponding to TOR and resuming of control. The difference represented the time in frames which was converted to seconds by dividing the difference by 30. By convention, readiness to drive is achieved when all the three actions are complete. Therefore, the maximum of the three response times was taken as the time to takeover (TOT). It is feasible (though not expected in the current study) that some drivers may resume all three actions in preparation for the TOR. In such instances, their data would have been excluded from the final analysis. However, no such instances were observed in the current study. To determine the influence of different factors on takeover time, the response time to motor readiness at TOR is modelled using a parametric duration model.

2.3.4 Post-drive Questionnaire

Finally, a questionnaire was used to collect driver demographics and investigate how frequently they used various, existing voice-assistants, such as Google, Siri, Alexa etc. Previous studies suggest that experience of using/adopting various in-car driving assistant systems (DAS) such as lane keep assist, adaptive cruise control, etc. are likely to influence the takeover time and acceptance of highly automated vehicles among drivers (Brookhuis et al., 2001; Molnar et al., 2018; Zeeb et al., 2015). In this study, it was assumed that such an acceptance and trust in the automated vehicle may encourage the drivers to completely relax during the period of automation. Similarly, any driver with prior experience of using various voice assistance technologies are likely to be more comfortable using VA compared to other drivers. Previously, Zeeb et al. (2015) used a 5-point Likert scale (1=no experience, 5=regular use) to rate the pre-user experience of DAS. Therefore, a similar scale was used in this study to collect the ratings on pre-user experience of drivers with various types of existing voice assistant technologies such as Google, Siri, Alexa etc. A higher rating would reflect higher exposure to such technology and likely to increase the acceptance of a system such as VA proposed in this study (Brookhuis et al., 2001; Molnar et al., 2018).

3 ANALYSIS AND RESULTS

The following section initially presents a summary of the dataset and participants' subjective responses to help contextualise results. This is followed by the comparative analysis of alertness measures (subjective ratings, visual behaviour) and situation awareness measures across the two drives. Finally, we introduce a parametric duration model of takeover time to study the effect of using VA, including driver characteristics collected through questionnaire and the visual behavior parameters as covariates in the model.

3.1 Dataset

Driver demographics and descriptive statistics are summarized in Table 3.

Table 3 Descriptive statistics of subjective data through questionnaires and visual indicators of sleep/fatigue

Categorical Variables (N=24)	**Categories**		**Frequency**	**Mean**	**SD (±)**
Gender	Male = 0		14		
	Female = 1		10		
Likeliness to sleep in an automated vehicle if it is allowed	very likely =4		6	2.63	1.498
	somewhat likely =3		9		
	neither likely nor unlikely = 2		2		
	somewhat unlikely = 1		2		
	very unlikely = 0 (reference)		5		
Continuous Variables (N=24)		**Min**	**Max**	**Mean**	**SD (±)**
Age (in years)		22	59	30.1	8.4
Average annual mileage (miles/year)		150	15000	4315.1	4114.1
No. of years of holding a valid driving licence		2	33	10.5	6.9
Average duration of sleep/day (in hours/ day)		5.00	9.00	7.2	0.9
Frequency of using VAs while in car (max.20)		4	14	6.8	3.2
Subjective Rating (N=24)		**Min**	**Max**	**Mean**	**SD (±)**
NASA TLX workload scale during automated drive	without VA	6	72	35.5	13.8
	with VA	8	64	41.0	12.0
Visual indicators of sleepiness/fatigue (N=23)					
Average eye blink duration during automation (ms)	without VA	233.5	700.0	388.2	120.8
	with VA	217.6	700.0	365.3	103.4
Average pupil diameter in mm during automation	without VA	1.7	5.4	3.6	0.8
	with VA	2.2	5.1	3.8	0.7
Average eye blink frequency during automation (blinks/s)	without VA	0.0	0.9	0.4	0.2
	with VA	0.0	0.8	0.5	0.2

3.1.1 Prior use of various voice assistance systems

The likeliness of using a voice assistant in future vehicles might depend on drivers' frequency and prior experience of using various types of existing voice-assistance systems, such as Google, Siri, Alexa etc. Most of the drivers already had some experience using voice-based assistants for either route navigation or as a music player (Figure 4), but did not use these to stimulate any conversations, for example, voice-based web search etc. In this study, individual responses to the frequency of using a voice assistant in the examples given (as listed in Figure 4 and as used in the conversations designed in this study) is rated on a Likert scale (1=never to 5=always) (Zeeb et al., 2015). These were summed to provide a single covariate indicating the frequency of using VAs rated on a linear scale varying from 1-20. The mean rating is thus summarized in Table 3.

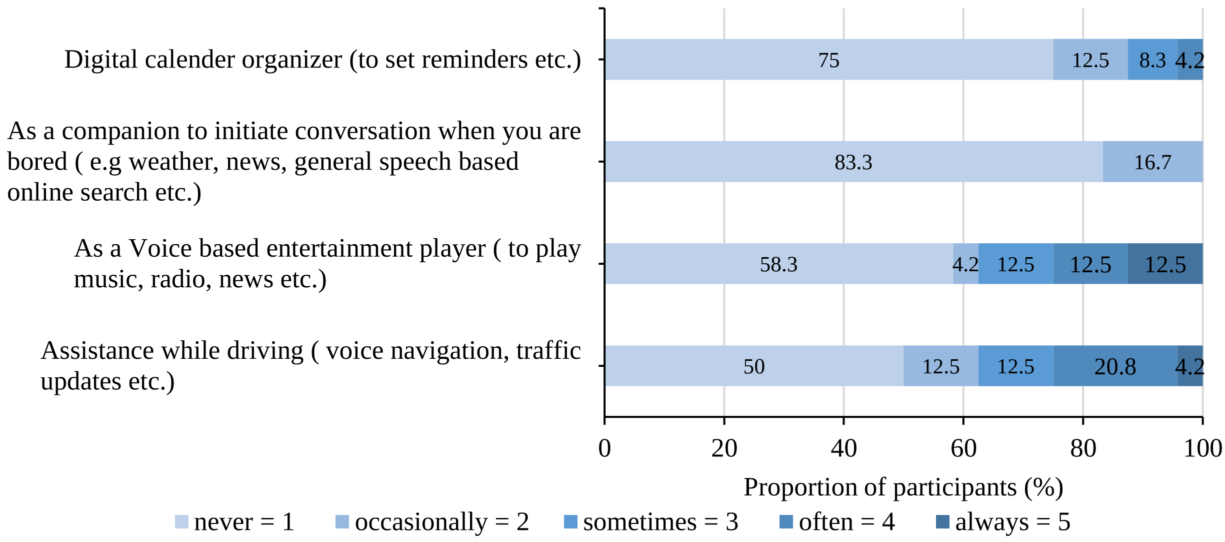


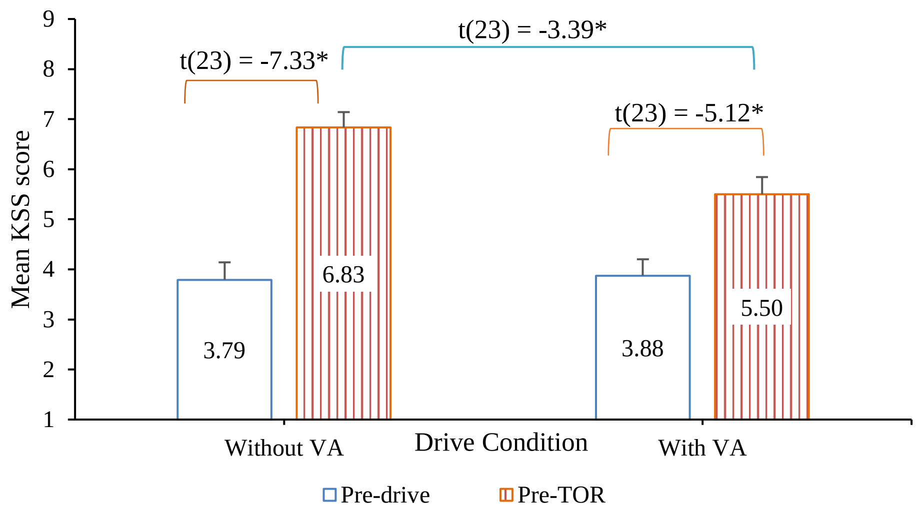
Figure 4 Frequency of using Voice Assistants such as Google, Alexa, Siri etc in different ways (proportion of participants in each response category is indicated corresponding to the specific use).

3.2 Alertness measures

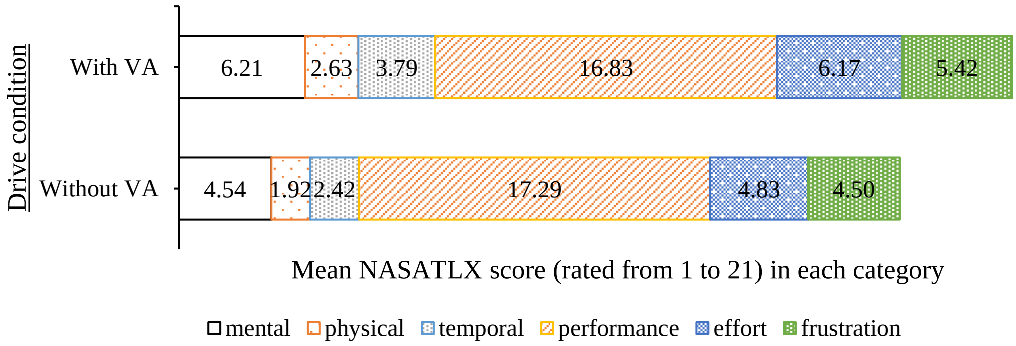
Paired samples t-tests and Wilcoxon signed rank tests were used to analyse the effect of interacting with VA on driver alertness towards the end of automation (i.e. just prior to TOR), compared with the drive absent of the VA.

3.2.1 Subjective alertness and workload

The mean KSS and combined NASA-TLX workload ratings (Figure 5), were compared using paired-samples t-tests. The t-tests showed a significant increase in mean KSS scores towards the end of automation (i.e. just prior to TOR) from pre-drive in both the drives (Figure 5a). The t-tests are indicated in Figure 5a. This indicates the fatiguing effects of automation, which persisted post-takeover as indicated by the KSS ratings, specifically in the absence of VA. However, the mean KSS rating just pre-TOR (i.e. post-automation) was significantly lower in the drive with VA ($t(23) = 3.391$, $p<0.005$) indicating higher alertness in presence of VA. Further, the combined NASA-TLX workload ratings were significantly higher at pre-TOR (i.e. just prior to TOR) when the drivers were accompanied by VA ($t(23) = -2.448$, $p<0.05$). Participants reported higher workload – suggesting higher alertness – during the drive with the VA, confirming that automation for long periods lowers cognitive workload and makes the drivers vulnerable to symptoms of passive fatigue.



a. Higher mean KSS scores (vertical bars represent standard deviation) post-automation just prior to TOR indicating loss of alertness (*represents significant values i.e. p<0.005)



b. Increase in mean NASATLX workload ratings during automation in the drive with VA

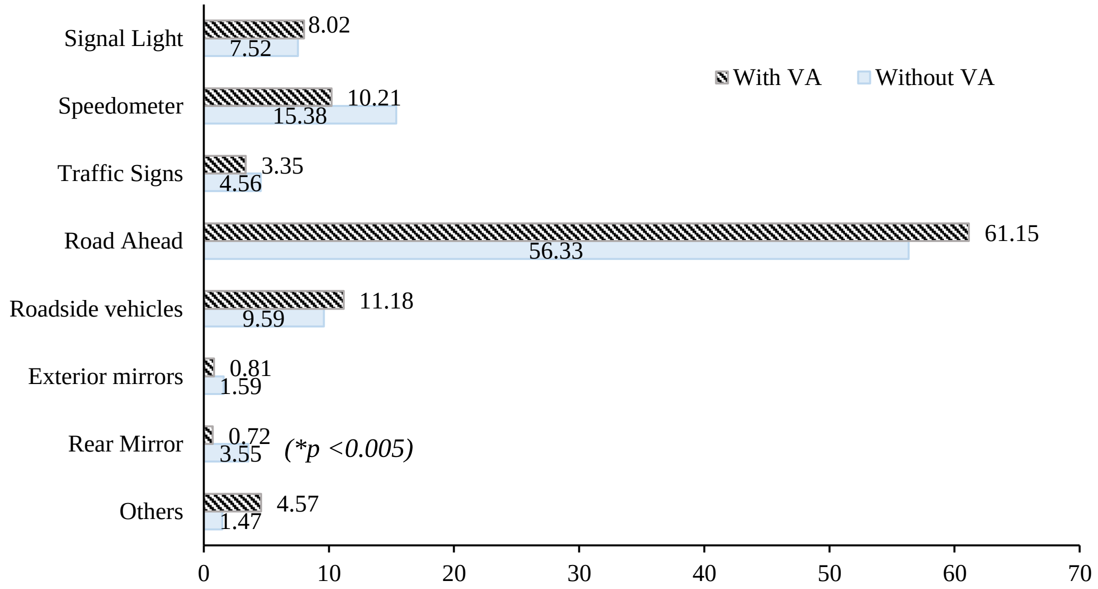
Figure 5 Mean KSS rating and NASATLX workload ratings compared across two drives

3.2.2 Objective alertness (eye-based) measures

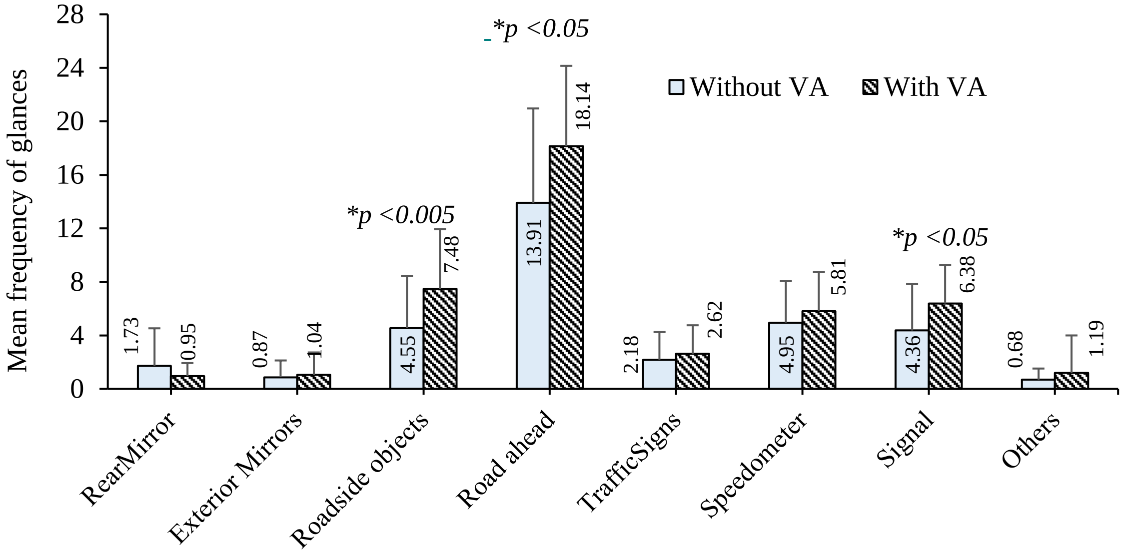
The visual characteristics data for one of the participants was unavailable due to technical malfunctioning of eye-tracking glasses during one of the drives. Thus, paired comparisons of visual indicators of fatigue during automation were conducted for the remaining 23 participants across the two drives. Among the eye blink duration (ms), frequency of blinking (blinks/s) and pupil diameter (mm), only average pupil diameter was found to be significantly different between the two drives ($t(21) = -2.263$, $p<0.05$). Larger pupil diameter (in mm) was observed during the drive with VA (Mean=3.77, SD = ±0.69) compared to drive without VA (Mean=3.58, SD = ±0.81), suggesting the former drivers were more alert.

3.3 Situation awareness near TOR

The AOI data are summarized in Figure 6 for the two drives. The proportion of time spent glancing at each AOI was calculated from the total glance duration of each participant. The time spent and glance frequency data for each AOI is then compared using a non-parametric Wilcoxon signed rank test across the two drives. Only four pairs showed significant differences as indicated in Figure 6.



a. Average proportion of total glance duration spent at defined AOIs (%)



b. Glance frequency at different AOIs

Figure 6 a. Proportion of time spent calculated from total glance duration and b. Glance frequency both averaged over all participants in each drive. The AOI statistics are compared using paired Wilcoxon signed rank test across two drives at takeover (* the pairs with significant differences in the two drives are indicated with their p-values)

Participants spent significantly more time glancing at the rear-view mirrors in the drive without VA (Mean= 3.55%, SD =±5.74%) compared to the drive with VA (Mean = 0.73%, SD = 0.85%). Comparison of glance frequency showed that drivers directed significantly more glances to the road in front, roadside objects (construction zone or parked vehicles on roadside) and upcoming traffic signal lights during the drive with VA, compared to the drive without VA, suggesting that they were more alert and engaged with the driving task when accompanied by the VA.

3.4 Modelling takeover time (TOT)

In this study, the voice assistant is intended to help the driver stay alert during automation and gain situation awareness at takeover. The results of the visual behavior and AOI statistics provide preliminary evidence that the drivers were more engaged with the road environment and more alert during the automated drive with VA which is likely to influence the takeover time. Therefore, a statistical modelling approach was adopted to quantify the contribution of these factors on takeover time. The parametric duration model, or survival analysis approach, is traditionally used in traffic accident research to study the response time of drivers to accidents incorporating the influence of other covariates (Choudhary and Velaga, 2017a; Haque and Washington, 2015, 2014; Mahajan and Velaga, 2020; Washington et al., 2003; Yadav and Velaga, 2019).

3.4.1 Parametric duration model

Parametric duration modelling is a probabilistic approach to analyse the conditional probability of the elapsed time until the event of interest, provided the event continues to time, t (Washington et al., 2003). In this study, the **event** is defined as "gaining motor readiness as shown by hands-on-wheel, feet on pedals and eyes on road" and the length of time to gain complete motor readiness in response to TOR is the **duration variable (T). 'T'** represents a continuous random variable with probability density function (PDF), f(t), given by equation (1) and cumulative distribution function (CDF) F(t) (Washington et al., 2003). F(t) is sometimes also referred to the failure condition, given by the probability of resuming manual driving before some specified time, t, in this study. Hence, the remaining probability of resuming manual control after the time 't' is called the **survivor function, S(t)** i.e. the probability of failing to resume manual driving before the construction zone appeared.

$$f(t) = \frac{dF(t)}{dt} = \frac{dP(T<t)}{dt} = \frac{d\,\{1-P(T\geq t)\}}{dt} = \frac{d\{1-S(t)\}}{dt} \qquad (1)$$

The **hazard function, h(t)** which is also called the instantaneous failure rate, gives the conditional probability that the event will occur between the time t and (t+dt) provided the event has continued for 't' or more duration (Haque and Washington, 2014; Washington et al., 2003).

$$\mathrm{h(t)} = \frac{\mathrm{f(t)}}{[1-\mathrm{F(t)}]} = \frac{\mathrm{f(t)}}{\mathrm{S(t)}} \qquad (2)$$

Parametric duration modelling is commonly adopted with proportional hazards (PH) assumption. The PH model is appropriate to use if the hazard ratio is constant over time for an individual. It assumes that the covariates act multiplicatively on some underlying baseline hazard function i.e. the explanatory variable only affects the chance of failure and not the failure time directly. Therefore, in this study the aim is to understand the underlying process of takeover or the survival time, which is affected by different covariates, rather to merely predict the takeover completion. Therefore, an alternative approach of accelerated failure time (AFT) model was used where the failure time is assumed to follow a distribution and the covariates directly influence the failure time without assuming a constant hazard ratio. In the baseline survivor function, all the exposure variables are kept at zero. An AFT model allows the covariates to rescale (accelerate) time directly in the baseline survivor function (Washington et al., 2003). Adopting an AFT approach allows us to capture the direct influence of an exposure variable on the mean survival time. Thus, the estimated parameters in the results directly show the effect of a covariate on mean response time. According to Washington et al. (2003), the natural log of duration variable, ln(T), is linearly related to the covariates:

$$\ln(\boldsymbol{T}) = \beta\boldsymbol{X} + \varepsilon \tag{3}$$

where 'X' is a set of covariates, β is vector of estimated parameters and ε = error term in the AFT model. The AFT model and conditional hazard function are given by:

$$S(t|\boldsymbol{X}) = S_0[t \cdot EXP(\beta\boldsymbol{X})], \tag{4}$$

$$\mathrm{h}(\mathrm{t}|\mathbf{X}) = \mathrm{h}_0[t \cdot EXP(\beta\boldsymbol{X})]\mathrm{EXP}(\beta\mathbf{X}) \tag{5}$$

where S_0 and h_0 are respectively the baseline survivor and hazard function.

To use the parametric duration model, we assume a distribution followed by a hazard function. Here, as the probability of completing the takeover is likely to increase over time showing a positive duration dependence. Thus, the Weibull distribution is suitable to model the takeover time data with monotone hazard rate (due to positive duration dependence event) that increases exponentially with time. The Weibull-distribution with scale-parameter (P > 0) and location-parameter (λ > 0) is given by (Washington et al., 2003):

$$f(\mathrm{t}) = \mathrm{h(t)EXP}[-(\lambda\mathrm{t})^{\mathrm{P}}]\,; if\ P > 1\ when\ hazard\ is\ monotonously\ increasing \tag{8}$$

In the Weibull duration model, the hazard function and survival function are expressed as:

$$\mathrm{h}(\boldsymbol{t}) = (\lambda\mathrm{P})\ (\lambda\mathrm{t})^{P-1} \tag{9}$$

$$\mathrm{S}(\boldsymbol{t}) = EXP\ (-\lambda\mathrm{t}^{P}) \tag{10}$$

The present study involved the modelling of takeover time at the prompt of a TOR after automation in two driving conditions. As discussed in the literature (section 1.2), all the variables listed in Table 3 were used as explanatory variables in the model except for visual behavior component. The visual behavior components were taken from the AOI analysis. Among all the components shown in Figure 6, checking traffic signs, checking mirrors and the road ahead are significant components of safe driving (Gold et al., 2013; Samuel et al., 2016; Vlakveld et al., 2018; Wright et al. 2016). Overall, the glances at various AOIs contribute towards situation awareness which are likely to influence the takeover time. The preliminary analysis shows a significant effect of VA on driver alertness, NASA-TLX workload rating and situation awareness. Therefore, a correlation analysis was conducted to identify the independent variables which were significantly associated with takeover time (Field, 2009; Pallant, 2010). The KSS rating, pupil diameter and NASA-TLX workload rating just prior to the TOR were found to be significantly correlated with the drive condition (with or without VA) (Appendix-1). Therefore, drive condition was retained as an explanatory variable in the model. Among the glance behavior metrics, the frequency of glances at different AOIs were found to be significantly correlated with their respective proportion of total glance duration (refer variables 4 to 18 in Appendix-1). The glance proportion at rear mirror, signal light and roadside vehicles were also found significantly correlated with the drive condition, road ahead and traffic signs, respectively (Appendix-1). Therefore, variables showing relatively higher correlations with the dependent variable i.e. takeover time were retained as explanatory variables in the model.

The study involved repeated observations across the two drives with the same participants – this can cause intra-group correlations and heterogeneities. The Weibull AFT model with gamma frailty (analogous to the random effects model) or clustered heterogeneity (standard error is adjusted for possible correlations due to repeated measures) can be used to avoid the effect of such heterogeneities and erroneous interpretation of parameter estimates in the model (Haque and Washington, 2014). Both the Weibull AFT models were developed using Stata SE-16 (at 95% significance level). The frailty parameter, theta ($\theta = 0$), indicated that the variance of gamma heterogeneity was negligible and insignificant ($\chi^2 = 0$, $p = 1.00$). The models were estimated by the standard maximum likelihood methods. The standard errors are robust to heteroscedasticity to address the auto-correlation problem due to repeated trials involving same individuals (Alan Agresti, 2002; Castillo-Manzano et al., 2016; Fitzharris et al., 2017; Mahajan and Velaga, 2020; Papke, 1996; Pylkkönen et al., 2015). Among all comparable models with the covariates (variables related to glance behavior, driver demographics, drive condition, workload and frequency of using voice assistants), the final model with clustered heterogeneity was chosen with minimum Akaike's information criteria (AIC) and Bayesian information criteria (BIC) values (Haque and Washington, 2014; Washington et al., 2003). The scale parameter $p = 4.1$ (>1) confirms that the survival rate is decreasing with time (i.e. hazard rate or probability of taking over before reaching the obstacle increased with time). Table 4 summarizes the estimated coefficients or exponential of coefficients (hazard ratio) which directly represents the relative change in survival time duration with unit increment in the covariates (StataCorp, 2013).

The model results provide an insight into the process of gaining situation awareness during a takeover, with parameters such as, duration of checking exterior mirrors, road in front, speedometer and traffic signs. The model shows that participants were likely to resume control 39% faster in the drive with VA compared to the other drive. Also, the model results show that female drivers are likely to take 1.21 times longer time to resume motor readiness as compared to their male counterparts. In addition, individuals who indicated that they frequently used different types of existing VAs are likely to take 8% less time to resume control. The model results also show that longer glances at different components of driving scene prior to the TOR will help in reducing the probability of delays in takeover. Driver age and willingness to sleep in automated vehicles did not influence the takeover time significantly.

Table 4 Weibull AFT (with clustered heterogeneity) model estimates with takeover time (TOT) as dependent variable

Variable	**Exp (B)**	**Coefficient (B)**	**Std. Error**	**z**	**p-value**
1. Drive condition (Without VA*)					
With VA	0.61	-0.5	0.203	-2.47	<0.05
2. Age	1.00	-0.004	0.006	-0.72	ns
3. Female (vs Male)	1.21	0.19	0.119	2.01	<0.05
4. Mileage	1.00	0.00	0.000	-3.23	0.001
5. Using VAs	0.92	-0.081	0.021	-3.76	<0.001
6. Sleeping in auto-car	1.00	0.003	0.035	-0.08	ns
%Glance duration (or %time spent)					
7. External Mirrors	0.90	-0.106	0.039	-2.71	<0.05
8. Road ahead	0.97	-0.03	0.008	-3.66	<0.001
9. Speedometer	0.96	-0.04	0.008	-4.47	<0.001
10. Traffic signs	0.88	-0.132	0.028	-4.78	<0.001
Intercept		5.27	0.909	5.8	<0.001
Log psuedolikelihood		-7.03	-	p<0.001	
Scale parameter-p		4.1	1.43		
theta		2.16E-08	0.00		
Model fitness comparison		N	df	AIC	BIC
• with clustered heterogeneity		24	12	38.07	52.2
• with gamma frailty		24	13	40.06	55.38

Note: *ns: not-significant; VA: Voice assistant*

To gain further insights into the model results, separate survival curves (Figure 7, 8 and 9) are plotted substituting parameter estimates in Table 4 corresponding to each covariate in Eq. (10). The probabilities were calculated for the two driving conditions (with and without VA) at different times. All other variables were either kept at their reference category (if categorical variable) or corresponding means were substituted (for continuous variables) (refer Table 3 and Figure 6).

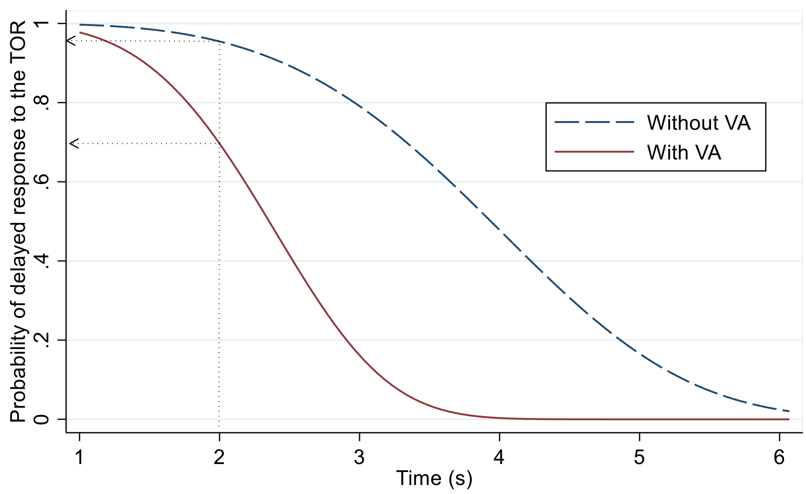


Figure 7 Survival curves for the two test conditions i.e., probability of not responding early to the TOR early

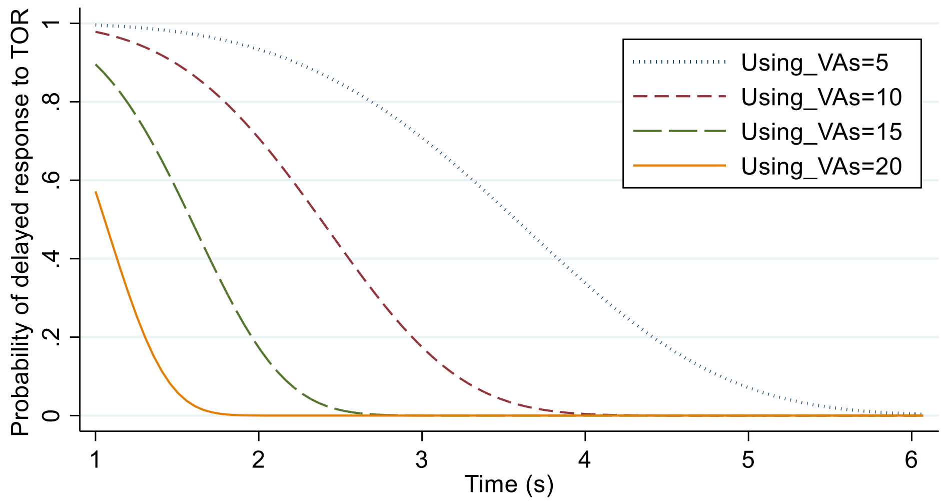


Figure 8 Reduction in probability of delayed takeover with higher frequency of using different types of voice assistants

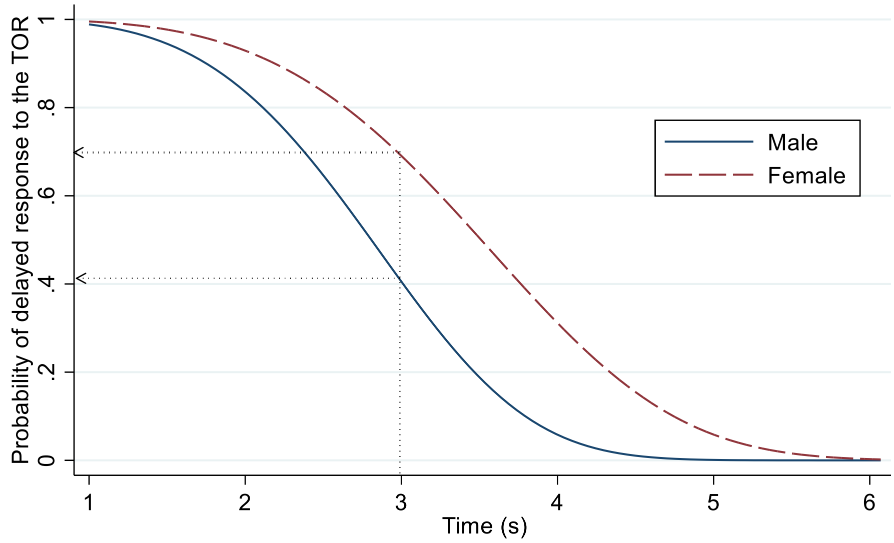


Figure 9 Effect of gender on probability of delayed response to takeover request

4 DISCUSSION

This study aimed to explore whether conversations with a digital voice assistant could help to keep a driver alert during automation and provide useful traffic information to assist the takeover process. The study findings are discussed as follows in response to the target questions raised in Section 1.4.1.

4.1 Effect of VA on driver alertness

Automation relieves the driver from the task of assessing the traffic scenario and taking required physical actions for driving, which results in low-workload conditions. Therefore, we hypothesized that providing a digital voice assistant will improve driver cognitive workload – to an *appropriate* level – through intermittent conversations. The increase in mean KSS ratings from pre-drive to post-automation (i.e. prior to TOR) indicates that automation significantly reduced driver's workload and caused passive fatigue. The paired comparisons show a significant increase in driver workload (as indicated by pre-TOR NASA-TLX workload ratings) due to the conversational exchanges with VA, suggesting improved alertness during automation. This is supported by both the subjective alertness measure i.e. lower mean KSS rating and objective measures, such as higher pupil diameter, during the drive with VA. Larger pupil-diameter indicates higher alertness during the drive with VA (Kapitaniak et al., 2015; Large et al., 2018; Wang and Xu, 2016). The significant correlation between alertness measures and NASA-TLX workload ratings also suggest that that the regular conversational interludes made by the VA interrupted the monotony of the automated drive increasing driver workload and thereby improving driver alertness. In combination with this, providing traffic–related conversations, such as informing drivers of the speed limit, upcoming intersections etc. kept the drivers more engaged with the driving environment. It is also noteworthy that none of the drivers were observed sleeping during the drive with VA, whereas six drivers had short episodes of "nodding off" when not accompanied by VA, and were notably startled by the takeover request.

4.2 Effect of VA on glance behavior

The AOI eye tracking analysis showed the differences in allocation of visual attention in response to the traffic feedback provided by VA near the TOR, supporting our second hypothesis. The VA pre-informed drivers about the new speed limit and an approaching pedestrian crossing, to engage them with the driving scene and act with caution. The paired comparison of AOI statistics between the two drives showed that for similar proportion of glance duration in the two drives, significantly higher number of glances were associated with checking exterior mirrors, concentrating on the road ahead and checking upcoming traffic signal during the drive with VA. The frequent glances at these specific objects also relate to the verbal message delivered by VA, during the one-minute prior to TOR, which shows that verbal cues can direct drivers' visual attention in the driving scene. For instance, information about the pedestrians and changing speed limit might have encouraged the drivers to confirm the information provided by VA by reconciling this with the driving scene. Secondly, this information might have led to increased mirror-checks and additional focus made by drivers on the road ahead to prepare themselves for any required action such as decelerating or braking at the traffic signal or to avoid colliding with pedestrians. Previous studies have also associated increased mirror checks with gaining situation awareness at takeover (Gold et al., 2018; Zeeb et al., 2016, 2015). Contrary to Vogelpohl et al. (2018), who reported a delay of 14-15s in the first glance to speedometer post-TOR in the absence of any traffic signs or a VA, the drivers in this study tended to shift their gaze immediately to the speedometer after the posting of a new speed limit sign. Such behavior was observed during both the drives in this study, irrespective of the fact that automation was engaged, and drivers were not engaged in manual driving. Therefore, the constant checking of speed limit and other traffic signs in both the drives might be an artefact of the simulated driving environment and explain why there was no difference in glances (duration or frequency) at the speedometer or traffic signs between both the drives. However, allocating glances to confirm the behavior of the automated vehicle in response to information obtained from traffic signs or by VA, may also suggest a lack of trust and acceptance by the drivers (both for automation and the digital assistant), although this may change over time, as drivers' experience with such systems increase (Brookhuis et al., 2001; Molnar et al., 2018; Zeeb et al., 2015).

Another interesting finding was the higher proportion of glance duration at the rear-view mirror in the absence of VA. During the one-minute period of AOI analysis, the driver was driving in the urban scenario, which included dynamic traffic surroundings, such as pedestrians, and unique buildings on the roadside. The need to frequently scan the rear-view mirrors arises during manoeuvres such as turning, changing lanes, merging, high traffic or during any hazard event requiring sudden braking or deceleration. Such cues were absent in the rear view and there was no following vehicle close to the participant's vehicle or any event which could have occupied their attention in the rear-view mirror for an extended period of time. Even though the frequency of glances at the rear mirror were similar in both the drives, however, the longer glances at the rear mirror in the drive without VA indicate a longer off-road time and disengagement from forward roadway. In the video recordings of experiments, three participants indicated that they were observing the vehicles or other objects in the scenario disappearing in the far view of rear mirrors out of curiosity. This was an artefact of the simulation software used. A few other participants also shared similar opinion about virtual driving scenario. However, as this was not included in the post drive questionnaire (please refer Mahajan et al. (2021) for more details on subjective analysis of post study feedback), therefore a precise count of such opinions was not recorded in the current study. Previous studies have associated checking the rear-view mirror as a positive step towards gaining situation awareness prior to take-over in hazard situations (Gold et al., 2018; Lu et al., 2017; Zeeb et al., 2015). However, NHTSA (2006) suggested that short off-road glances (<2s) at rear-view mirrors or familiar traffic signs are sufficient to scan the driving scene to reduce the risk. Therefore, in the drive without VA i.e., without any vocal alerts to redirect their gaze back on the road, drivers may have spent longer glances checking the virtual environment displayed in the rear-view mirror.

4.3 VA and the takeover process

4.3.1 Influence of VA on takeover time

In order to improve our overall understanding of the takeover process, and more specifically, how VA can directly (driving with VA) or indirectly (through changing gaze behavior) influence drivers' response to a TOR, a parametric duration modelling approach was adopted. The parametric duration model shows conclusively that intervention with the VA helped drivers to gain readiness to resume manual controls quicker compared to the other condition (without VA). The survival graph in Figure 7 shows the probability of a delayed response to takeover times is much higher, relatively, during the drive without VA compared to the drive with VA. A timely takeover was defined with respect to a construction zone in roadside. When asked to take over control, drivers were expected to check the surrounding environment prior to the construction zone and resume driving controls to avoid the risk of the car veering into the construction zone. During the drive with VA, drivers were more alert, and the evidence suggests that they noticed the construction zone following the takeover request. Moreover, they had prior information of an intersection signal ahead and an upcoming change in the speed limit. The intention in providing this information was to prepare them for driving, even before the takeover request was issued. However, during the drive without the VA, drivers were not only fatigued and sleepy, but had been provided with no information relating to impending traffic signals or changing speed limits etc. Therefore, it is suspected that the process of becoming alert and building situation awareness would have been responsible for delaying the takeover process during this drive.

4.3.2 Influence of gaze behavior on takeover time

The conversation relating to traffic and the road environment provided information which appears to have influenced drivers' gaze behavior, encouraging them to scan the driving scene, including traffic signs, their speedometer, mirrors etc., thereby maintaining their situation awareness and supporting a quicker response to the takeover request. Zeeb et al. (2015) also claimed that gaze behavior is a significant indicator of cognitive process at the TOR, and this determines the takeover time. A few previous studies associated the increase in glance duration at traffic signs, mirrors etc. indicating higher processing time to understand and respond to stimuli such as traffic signs (Babi et al., 2020; Martens and Fox, 2007). Although there was no significant difference in driver glance duration at most of the AOIs during the one-minute leading to the TOR, across the two drives, however, the glance frequencies (at road ahead, mirror checks and signal) were relative higher in the drive with VA. Therefore, providing drivers with access to such information even before the TOR (as provided by VA in this study) may reduce their processing time at TOR. This may explain why the increase in glance durations at the road in front, exterior mirrors, speedometer and traffic signs in the model support the positive influence of scanning various components of driving scene prior to the TOR, thereby reducing the probability of any delays in responding to TOR.

4.3.3 Pre-user experience of VAs and takeover time

It was apparent that drivers who indicated that they frequently used other voice-based digital assistants felt more comfortable using VA, and this may have encouraged them to engage more in conversations (which might also be in terms of attentive listening). Despite the lack of exposure and availability of various advanced driver assistance technologies to many of the participants, they expressed their intention to use similar voice-based driver assistant systems in the future. The results of the study suggest that this is likely to have a positive impact on factors such as alertness as was assumed by Zeeb et al. (2015). This is also shown by the survival curves plotted in Figure 8, using hypothetical rating on pre-user experience of various voice assistants by the drivers. As suggested in Figure 4 and section 3.1.1, a hypothetical increase in rating from 5 (i.e., very rare use of any voice assistant) to 20 (very often or regularly using different types of voice assistants) represents an increase in adoption of such technology. This suggests that higher exposure to using such systems could potentially increase their effectiveness in assisting the drivers to takeover controls after automation even further.

4.3.4 Other factors influencing takeover time

According to the model results, female drivers are likely to take longer to take over control (or to demonstrate motor readiness) than male drivers (Figure 9). Such a finding is interesting and could reflect a more cautious approach amongst female drivers, who may spend more time exploring and assessing the driving scene at the TOR. Similar results were reported by (Gomez et al., 2019; Lings, 1991). Among the various non-driving activities that drivers could perform during automation, sleeping might also be a voluntary action rather than just induced by the automation (Wan and Wu, 2018). In this study, 15 out of 24 drivers responded that they would likely sleep in an automated vehicle, if it were possible (Table 3). Nevertheless, willingness to sleep also suggests high trust and acceptance in the technology. According to the model results, such willingness to sleep in an automated vehicle did not affect the probability of delays in takeover time (Table 4).

5 CONCLUSION

Extended periods of highly automated driving can disengage drivers from the driving task and reduce their alertness. The study explored the use of a digital voice assistant (VA) which could act as a driving coach/assistant during periods of highly automated driving. The study findings highlight the following benefits of adopting a futuristic digital voice assistant in partially automated vehicles:

i. The VA could counter the effects of passive fatigue during highly automated drive, supporting our hypotheses.

ii. Traffic-related information delivered by the VA can direct driver's visual attention to traffic signs, exterior mirrors or road-ahead and assist in building situation awareness required during a takeover event.

iii. The takeover time reduced significantly during the drive with VA, with appropriate situation awareness. The pre-allocation of visual attention to different components of driving environment due to verbal messages from VA also contributed to reducing the takeover time.

Further, the parametric model of takeover time highlighted, amongst other findings, gender-based differences in takeover time of drivers. This approach can be extended to find the effects of other relevant variables such as driver age, exposure to various in-car driver assistance systems, which may influence drivers' acceptance and trust towards automated vehicles as well as towards systems like VA. The study also revealed that in the absence of technological interventions, drivers may choose to sleep in highly automated vehicles. This highlights the need for driver-assistance systems similar to the one proposed here to ensure that drivers are prepared to takeover control with appropriate alertness and awareness of the current traffic situation.

In practice, for a voice assistant to have the capacity to receive and disseminate such information may depend on the efficacy and availability of intelligent transportation systems. Also, the VA would need to be coordinated with the other driving assistance systems in the vehicle such as controlling speed, drifting lanes etc. to follow the final decisions of either driver/VA. The short, intermittent conversations of 30-60s, conducted at an approximate interval of 3 minutes, were found to be effective in this study. The conversation also included a few general conversation topics such as event reminders, entertainment etc. which may promote frequent use of such technology to build trust and adaptation of users towards such technology to make it successful. Future research could investigate the detailed interface design requirements and enhancing the connectivity and interoperability of such systems. As a final note, the study demonstrated the positive effects of using a voice assistant providing regular conversation during periods of automation. However, such effects are likely to be transient, and therefore more research is required to investigate the lasting effects of such interventions.

5.1 Limitations and Future Scope

Finally, it is important to recognise the limitations of our study which should be taken into account before generalizing from the results. Firstly, sample size was limited. However, as indicated in Table 1, many related studies have used a small sample size previously (Gonçalves et al., 2016; Vlakveld et al., 2018; Vogelpohl et al., 2018; Wu et al., 2019). Also, due to the smaller proportion of each gender group, the gender-based findings should be interpreted with caution. Further, excluding the participants based on various criteria such as excessive day-time sleepiness symptoms or sleep disorders may limit the generalization of findings over the entire population. The sleep duration or consumption of caffeine, mints etc. prior to the study were not supervised or monitored by the experimenter in this study. The minor standard deviations (0.3 to 0.35) in the mean KSS ratings measured pre-drive or during the drives suggest that the participants possibly did not violate these instructions. However, it might be preferable to monitor such conditions to validate results. The study included a small variety of short conversation topics mostly focused on providing traffic information (to avoid misinterpretations/miscommunication in lengthy conversations). Alternate studies may explore the engaging or distracting effects of involving follow up conversations on one single topic.

Building on the results, further research could explore different strategies for allocating visual attention, for example, identifying where the voice assistant should direct drivers' attention. Future studies can also observe the gradual change in alertness and visual behavior during the entire period of automation to gain further insights into effectiveness of VA. Another area worthy of further investigation is that of driver age. Younger drivers are expected to be more tech-savvy and therefore more likely to use voice-based technologies than older drivers – who may subsequently not receive the benefits highlighted in the study.

ACKNOWLEDGEMENTS

This research was conducted through the financial support from the Commonwealth Scholarship Commission (CSC) of UK under the split-site PhD research grant, 2018-2019 (Ref: INCN-2018-92). The contents of this manuscript are the sole responsibility of the authors of this paper and can in no way be taken to reflect the views of the CSC.

REFERENCES

Åkerstedt, T., Gillberg, M., 1990. Subjective and Objective Sleepiness in the Active Individual. Int. J. Neurosci. 52, 29–37. https://doi.org/10.3109/00207459008994241

Alan Agresti, 2002. Categorical Data Analysis, Wiley-Interscience.

Antrobus, V., Burnett, G., Large, D.R., 2018. ' Trust me – I ' m AutoCAB ': Using natural language interfaces to improve the trust and acceptance of level 4 / 5 autonomous vehicles 13–14.

Atchley, P., Chan, M., Gregersen, S., 2013. A Strategically Timed Verbal Task Improves Performance and Neurophysiological Alertness During Fatiguing Drives. Hum. Factors 56, 453–462. https://doi.org/10.1177/0018720813500305

Babi, Dario, Dijani, H., Jakob, L., Babi, Darko, Garcia-garzon, E., 2020. Driver eye movements in relation to unfamiliar traffic signs : An eye tracking study. Appl. Ergon. 89. https://doi.org/10.1016/j.apergo.2020.103191

Brookhuis, K.A., Waard, D. De, Janssen, W.H., 2001. Behavioural impacts of Advanced Driver Assistance Systems – an overview. Eur. J. Transp. Infrastruct. Res. 1.

Campagne, A., Pebayle, T., Muzet, A., 2005. Oculomotor changes due to road events during prolonged monotonous simulated driving. Biol. Psychol. 68, 353–368. https://doi.org/10.1016/j.biopsycho.2004.05.003

Castillo-Manzano, J.I., Castro-Nuño, M., Fageda, X., 2016. Exploring the relationship between truck load capacity and traffic accidents in the European Union. Transp. Res. Part E Logist. Transp. Rev. 88, 94–109. https://doi.org/10.1016/j.tre.2016.02.003

Centofanti, S.A., Dorrian, J., Hilditch, C.J., Banks, S., 2017. Do night naps impact driving performance and daytime recovery sleep? Accid. Anal. Prev. 99, 416–421. https://doi.org/10.1016/j.aap.2015.11.009

Choudhary, P., Velaga, N.R., 2017a. Modelling driver distraction effects due to mobile phone use on reaction time. Transp. Res. Part C Emerg. Technol. 77, 351–365. https://doi.org/10.1016/j.trc.2017.02.007

Choudhary, P., Velaga, N.R., 2017b. Mobile phone use during driving: effects on speed reduction and effectiveness of compensatory behaviour. Accid. Anal. Prev. J. 106, 370–378. https://doi.org/10.1016/j.aap.2017.06.021

Choudhary, P., Velaga, N.R., 2017c. Analysis of vehicle-based lateral performance measures during distracted driving due to phone use. Transp. Res. Part F Traffic Psychol. Behav. 44, 120–133. https://doi.org/10.1016/j.trf.2016.11.002

Dozza, M., 2013. What factors influence drivers ' response time for evasive maneuvers in real traffic ? Accid. Anal. Prev. 58, 299–308. https://doi.org/10.1016/j.aap.2012.06.003

Drews, F.A., Pasupathi, M., Strayer, D.L., 2008. Passenger and Cell Phone Conversations in Simulated Driving. J. Exp. Psychol. Appl. 14, 392–400. https://doi.org/10.1037/a0013119

Field, A., 2009. Discovering Statistics using SPSS, Third. ed. SAGE Publications Ltd.

Fitzharris, M., Liu, S., Stephens, A.N., Lenné, M.G., 2017. The relative importance of real-time in-cab and external feedback in managing fatigue in real-world commercial transport operations. Traffic Inj. Prev. 18, S71–S78. https://doi.org/10.1080/15389588.2017.1306855

Gold, C., Damböck, D., Lorenz, L., Bengler, K., 2013. Take over! How long does it take to get the driver back into the loop? Proc. Hum. Factors Ergon. Soc. 57, 1938–1942. https://doi.org/10.1177/1541931213571433

Gold, C., Happee, R., Bengler, K., 2018. Modeling take-over performance in level 3 conditionally automated vehicles. Accid. Anal. Prev. 116, 3–13. https://doi.org/10.1016/j.aap.2017.11.009

Gomez, P., Gunten, A. Von, Danuser, B., 2019. Eye gaze behavior during affective picture viewing : Effects of motivational significance , gender , age , and repeated exposure. Biol. Psychol. 146, 107713. https://doi.org/10.1016/j.biopsycho.2019.06.001

Gonçalves, J., Happee, R., Bengler, K., 2016. Drowsiness in conditional automation: Proneness, diagnosis and driving performance effects. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 31, 873–878. https://doi.org/10.1109/ITSC.2016.7795658

Haque, M., Washington, S., 2015. The impact of mobile phone distraction on the braking behaviour of young drivers : A hazard-based duration model. Transp. Res. Part C 50, 13–27. https://doi.org/10.1016/j.trc.2014.07.011

Haque, M., Washington, S., 2014. A parametric duration model of the reaction times of drivers distracted by mobile phone conversations. Accid. Anal. Prev. 62, 42–53. https://doi.org/10.1016/j.aap.2013.09.010

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Adv. Psychol. 52, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hilditch, C.J., Dorrian, J., Centofanti, S.A., Van Dongen, H.P., Banks, S., 2017. Sleep inertia associated with a 10-min nap before the commute home following a night shift: A laboratory simulation study. Accid. Anal. Prev. 99, 411–415. https://doi.org/10.1016/j.aap.2015.11.010

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., Dement, W.C., 1973. Quantification of Sleepiness: A New Approach. Psychophysiology 10, 431–436. https://doi.org/10.1111/j.1469-8986.1973.tb00801.x

Jamson, A.H., Merat, N., Carsten, O.M.J., Lai, F.C.H., 2013. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. Transp. Res. Part C 30, 116–125. https://doi.org/10.1016/j.trc.2013.02.008

Kapitaniak, B., Walczak, M., Kosobudzki, M., Jóźwiak, Z., Bortkiewicz, A., 2015. Application of eye-tracking in drivers testing: A review of research. Int. J. Occup. Med. Environ. Health 28, 941–954. https://doi.org/10.13075/ijomeh.1896.00317

Kohl, J., Gross, A., Henning, M., Baumgarten, T., 2020. Driver glance behavior towards displayed images on in-vehicle information systems under real driving conditions. Transp. Res. Part F Psychol. Behav. 70, 163–174. https://doi.org/10.1016/j.trf.2020.01.017

Körber, M., Cingel, A., Zimmermann, M., Bengler, K., 2015. Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving, in: Procedia Manufacturing. pp. 2403–2409. https://doi.org/10.1016/j.promfg.2015.07.499

Lal, S.K.L., Craig, A., 2001. A critical review of the psychophysiology of driver fatigue. Biol. Psychol. 55, 173–194. https://doi.org/10.1016/S0301-0511(00)00085-5

Large, D.R., Burnett, G., Antrobus, V., Skrypchuk, L., 2018. Driven to discussion: engaging drivers in conversation with a digital assistant as a countermeasure to passive task-related fatigue. IET Intell. Transp. Syst. 12, 420–426. https://doi.org/10.1049/iet-its.2017.0201

Large, D.R., Burnett, G., Anyasodo, B., Skrypchuk, L., 2016. Assessing Cognitive Demand during Natural Language Interactions with a Digital Driving Assistant. Proc. 8th Int. Conf. Automot. User Interfaces Interact. Veh. Appl. - Automotive'UI 16 67–74. https://doi.org/10.1145/3003715.3005408

Large, D.R., Clark, L., Quandt, A., Burnett, G., Skrypchuk, L., 2017. Steering the conversation : A linguistic exploration of natural language interactions with a digital assistant during simulated driving. Appl. Ergon. 63, 53–61. https://doi.org/10.1016/j.apergo.2017.04.003

Lings, S., 1991. Assessing driving capability: A method for individual testing The significance of paraparesis inferior studied in a controlled experiment. Appl. Ergon. 22, 75–84. https://doi.org/10.1016/0003-6870(91)90305-2

Lomelí, H.A., Pérez-Olmos, I., Talero-Gutiérrez, C., Moreno, C.B., González-Reyes, R., Palacios, L., De La Peña, F., Muñoz-Delgado, J., 2008. Sleep evaluation scales and questionnaires: A review. Actas Esp. Psiquiatr. 36, 50–59.

Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., Merat, N., 2017. Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. Accid. Anal. Prev. 108, 9–18. https://doi.org/10.1016/j.aap.2017.08.011

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. https://doi.org/10.1016/j.apergo.2016.12.003

M.W. Johns, 1990. ESS Test 1990-97 1990.

Mahajan, K., Large, D.R., Burnett, G., Velaga, N.R., 2021. Exploring the effectiveness of a voice assistant to maintain alertness in partially automated vehicles. Traffic Inj. Prev. https://doi.org/10.1080/15389588.2021.1904138

Mahajan, K., Velaga, N.R., 2020. Effects of Partial Sleep Deprivation on Braking Response of Drivers in Hazard Scenarios. Accid. Anal. Prev. 142, 1–22. https://doi.org/10.1016/j.aap.2020.105545

Martens, M.H., Fox, M.R.J., 2007. Do familiarity and expectations change perception ? Drivers ' glances and response to changes. Transp. Res. Part F Traffic Psychol. Behav. 10, 476–492. https://doi.org/10.1016/j.trf.2007.05.003

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., Warm, J.S., 2002. Fundamental Dimensions of Subjective State in Performance Settings : Task Engagement, Distress, and Worry. Emotion 2, 315–340. https://doi.org/10.1037/1528-3542.2.4.315

Matthews, G., Desmond, P., Neubauar, C., Hancock, P.A., 2012. The Handbook of Operator Fatigue, in: Matthews, G., Desmond, P., Neubauar, C., Hancock, P.A. (Eds.), The Handbook of Operator Fatigue. CRC Press. https://doi.org/10.1201/9781315557366-12

Matthews, G., Neubauer, C., Saxby, D.J., Wohleber, R.W., Lin, J., 2018. Dangerous intersections? A review of studies of fatigue and distraction in the automated vehicle. Accid. Anal. Prev. 0–1. https://doi.org/10.1016/j.aap.2018.04.004

Matthews, G., Szalma, J., Panganiban, A.R., Neubauer, C., Warm, J.S., 2013. Profiling task stress with the dundee state questionnaire, Psychology of Stress: New Research.

May, J.F., Baldwin, C.L., 2009. Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transp. Res. Part F Traffic Psychol. Behav. 12, 218–224. https://doi.org/10.1016/j.trf.2008.11.005

Mcdonald, A., Alambeigi, H., 2019. Toward Computational Simulations of Behavior During Automated Driving Takeovers : A Review of the Empirical and Modeling Literatures. J. Hum. Factors Ergon. Soc. 61, 642–688. https://doi.org/10.1177/0018720819829572

Mehmood, A., Easa, S.M., 2009. Modeling Reaction Time in Car-Following Behaviour Based on Human Factors. Int. Sch. Sci. Res. Innov. 3, 325–333.

Merat, N., Jamson, A.H., Lai, F.C.H., Daly, M., Carsten, O.M.J., 2014. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. Transp. Res. Part F Traffic Psychol. Behav. 27, 274–282. https://doi.org/10.1016/j.trf.2014.09.005

Miletin, M.S., Hanly, P.J., 2003. Measurement properties of the Epworth sleepiness scale. Sleep Med. 4, 195–199. https://doi.org/10.1016/S1389-9457(03)00031-5

Molnar, L.J., Ryan, L.H., Pradhan, A.K., Eby, D.W., St, R.M., Zakrajsek, J.S., 2018. Understanding trust and acceptance of automated vehicles : An exploratory simulator study of transfer of control between automated and manual driving. Transp. Res. Part F Psychol. Behav. 58, 319–328. https://doi.org/10.1016/j.trf.2018.06.004

Neubauer, C., Matthews, G., Langheim, L., Saxby, D., 2012. Fatigue and Voluntary Utilization of Automation in Simulated Driving. Hum. Factors J. Hum. Factors Ergon. Soc. 54, 734–746. https://doi.org/10.1177/0018720811423261

Neubauer, C., Matthews, G., Saxby, D., 2014. Fatigue in the Automated Vehicle. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 58, 2053–2057. https://doi.org/10.1177/1541931214581432

NHTSA, 2006. The Impact of Driver Inattention On Near Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. US Dep. Transp. 226. https://doi.org/DOT HS 810 594

Noble, A.M., Miles, M., Perez, M.A., Guo, F., Klauer, S.G., 2021. Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. Accid. Anal. Prev. 151, 105959. https://doi.org/10.1016/j.aap.2020.105959

Otmani, S., Pebayle, T., Roge, J., Muzet, A., 2005. Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. Physiol. Behav. 84, 715–724. https://doi.org/10.1016/j.physbeh.2005.02.021

Pallant, J., 2010. Survival Manual, McGraw-Hill Education.

Papke, L.E., 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. J. Appl. Econom. 11, 619–632. https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1

Philip, P., Taillard, J., Léger, D., Diefenbach, K., Akerstedt, T., Bioulac, B., Guilleminault, C., Philip, P., Taillard, J., Léger, D., Diefenbach, K., Akerstedt, T., Bioulac, B., Guilleminault, C., 2002. Work and rest sleep schedules of 227 European truck drivers. Sleep Med. 3, 507–511. https://doi.org/10.1016/S1389-9457(02)00138-7

Pylkkönen, M., Sihvola, M., Hyvärinen, H.K., Puttonen, S., Hublin, C., Sallinen, M., 2015. Sleepiness, sleep, and use of sleepiness countermeasures in shift-working long-haul truck drivers. Accid. Anal. Prev. 80, 201–210. https://doi.org/10.1016/j.aap.2015.03.031

SAE International, 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, SURFACE VEHICLE RECOMMENDED PRACTICE (SAE open platform).

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. J. Transp. Res. Board 2602, 115–120. https://doi.org/10.3141/2602-14

Sarkies, M.N., Skinner, E.H., Bowles, K., Morris, M.E., Williams, C., Brien, L.O., Bardoel, A., Martin, J., Holland, A.E., Carey, L., White, J., Haines, T.P., 2019. A novel counterbalanced implementation study design : methodological description and application to implementation research. Implement. Sci. 1–11. https://doi.org/10.1186/s13012019-0896-0

Saxby, D.J., Matthews, G., Neubauer, C., 2017. The relationship between cell phone use and management of driver fatigue: It’s complicated. J. Safety Res. 61, 129–140. https://doi.org/10.1016/j.jsr.2017.02.016

Saxby, D.J., Matthews, G., Warm, J.S., Hitchcock, E.M., Neubauer, C., 2013. Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts. J. Exp. Psychol. Appl. 19, 287–300. https://doi.org/10.1037/a0034386

Scullin, M.K., Bliwise, D.L., 2015. Sleep, Cognition, and Normal Aging. Perspect. Psychol. Sci. 10, 97–137. https://doi.org/10.1177/1745691614556680

Sheridan, B.T.B., Parasuraman, R., 2002. Human-Automation Interaction, in: Reviews of Human Factors and Ergonomics. pp. 89–129.

Shinar, D., Schechtman, E., Compton, R., 2001. Self-reports of safe driving behaviors in relationship to sex, age, education and income in the US adult driving population. Accid. Anal. Prev. 33, 111–116. https://doi.org/10.1016/S0001-4575(00)00021-X

Simons-Morton, B.G., Guo, F., Klauer, S.G., Ehsani, J.P., Pradhan, A.K., 2014. Keep your eyes on the road: Young driver crash risk increases according to duration of distraction. J. Adolesc. Heal. 54, S61–S67. https://doi.org/10.1016/j.jadohealth.2013.11.021

Smith, S.S., Horswill, M.S., Chambers, B., Wetton, M., 2009. Hazard perception in novice and experienced drivers: The effects of sleepiness. Accid. Anal. Prev. 41, 729–733. https://doi.org/10.1016/j.aap.2009.03.016

Starkey, N.J., Charlton, S.G., Malhotra, N., Lehtonen, E., 2020. Drivers ' response to speed warnings provided by a smart phone app. Transp. Res. Part C Emerg. Technol. 110, 209–221. https://doi.org/10.1016/j.trc.2019.11.020

StataCorp, 2013. Stata Survival Analysis and Epidemiological Tables, Release 13. ed, Stata Reference Manual. Stata Press, College Station, Texas.

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. https://doi.org/10.1016/j.trf.2018.07.025

Vogelpohl, T., Kühn, M., Hummel, T., Vollrath, M., 2018. Asleep at the automated wheel — Sleepiness and fatigue during highly automated driving. Accid. Anal. Prev. 0–1. https://doi.org/10.1016/j.aap.2018.03.013

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. Human-Machine Syst. PP, 1–10. https://doi.org/10.1109/THMS.2018.2844251

Wang, X., Xu, C., 2016. Driver drowsiness detection based on non-intrusive metrics considering individual specifics. Accid. Anal. Prev. 95, 350–357. https://doi.org/10.1016/j.aap.2015.09.002

Warshawsky-livne, L., Shinar, D., 2002. Effects of uncertainty , transmission type , driver age and gender on brake reaction and movement time. J. Safety Res. 33, 117–128.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC, United States of America.

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. Hum. Factors Ergon. Soc. 270–273. https://doi.org/10.1177/1541931213601062

Wu, Y., Kihara, K., Takeda, Y., Sato, T., Akamatsu, M., Kitazaki, S., 2019. Effects of scheduled manual driving on drowsiness and response to take over request: A simulator study towards understanding drivers in automated driving. Accid. Anal. Prev. 124, 202–209. https://doi.org/10.1016/j.aap.2019.01.013

Yadav, A.K., Velaga, N.R., 2019. Modelling the relationship between different Blood Alcohol Concentrations and reaction time of young and mature drivers. Transp. Res. Part F Psychol. Behav. 64, 227–245. https://doi.org/10.1016/j.trf.2019.05.011

Young, M.S., Stanton, N.A., 2002. Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. Hum. Factors 44, 365–375. https://doi.org/10.1518/0018720024497709

Young, R.A., 2013. Drowsy Driving Increases Severity of Safety-Critical Events and Is Decreased by Cell Phone Conversation. Proc. 3rd Int. Conf. Driv. Distraction Ina. 1–24.

Zeeb, K., Buchner, A., Schrauf, M., 2016. Is take-over time all that matters? the impact of visual-cognitive load on driver take-over quality after conditionally automated driving. Accid. Anal. Prev. 92, 230–239. https://doi.org/10.1016/j.aap.2016.04.002

Zeeb, K., Buchner, A., Schrauf, M., 2015. What determines the take-over time? An integrated model approach of driver take-over after automated driving. Accid. Anal. Prev. 78, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

Zhang, B., Winter, J. De, Varotto, S., Happee, R., Martens, M., 2019. Determinants of take-over time from automated driving : A meta-analysis of 129 studies. Transp. Res. Part F Traffic Psychol. Behav. 64, 285–307. https://doi.org/10.1016/j.trf.2019.04.020

Zhang, H., Wu, C., Yan, X., Qiu, T.Z., 2016. The effect of fatigue driving on car following behavior. Transp. Res. Part F Traffic Psychol. Behav. 43, 80–89. https://doi.org/10.1016/j.trf.2016.06.017

Appendix -1. Correlation coefficients between potential independent variables and dependent variable

Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
1. Test Condition	1																				
2. NASATLX rating Pre-TOR	0.314*	1																			
3. KSS score Pre-TOR	-.376**	-.318*	1																		
Glance frequency at:																					
4. G_RearMirror	--	--	0.343*	1																	
5. G_ExteriorMirrors	--	--	--	0.343*	1																
6. G_Roadside	.338*	--	-.318*	.305*	0.314*	1															
7. G_Road ahead	.314*	--	-.302*	--	--	.655**	1														
8. G_TrafficSigns	--	--	--	--	--	.713**	.558**	1													
9. G_Speedometer	--	--	--	--	--	--	.594**	.427**	1												
10. G_Signal	.306*	--	--	--	--	--	.498**	--	--	1											
11. G_Other	--	--	-.304*	--	--	--	--	--	--	--	1										
% Total glance duration at:																					
12. RearMirror	-.328*	--	--	.774**	--	--	--	--	--	--	--	1									
13. ExtMirrors	--	--	--	.369*	.675**	--	--	--	--	-.307*	--	0.314*	1								
14. Roadside	--	--	--	.417**	--	.787**	.373*	.772**	--	--	--	--	--	1							
15. Road ahead	--	0.343*	--	-.476**	--	--	--	-.466**	-.396**	--	--	-.441**	--	-.552**	1						
16. TrafficSigns	--	--	--	0.343*	--	.389**	-0.58**	.774**	0.419*	--	--	--	--	.683**	--	1					
17. Speedometer	--	--	--	--	--	--	--	--	-0.641**	--	--	.344*	--	--	--	--	1				

18. Signal	--	--	--	--	--	--	--	--	--	.671**	--	--	--	--	-.413**	--	--	1			
19. Others	--	--	--	--	--	--	--	--	--	--	.952**	--	--	--	-0.343*	--	--	--	1		
20. Pupil diameter Right (mm)	0.228*	-0.301*	-0.343*	0.343*	--	--	--	--	--	--	--	--	--	--	--	-0.343*	--	--	--	1	
21. Takeover time (TOT)	-0.305*	-0.261*	-0.206	--	-0.252*	-0.221*	0.201*	--	--	--	--	--	-0.316*	--	-0.307*	-0.313*	-0.261*	-0.215*	--	--	1

'*' coefficients significant at 0.05 level; '**' coefficients significant at 0.01 level; '--' not significant; 'G_XX' represents the frequency of glances at XX component of driving scene.