Exploring the Benefits of Conversing with a Digital Voice Assistant during Automated Driving: A Parametric Duration Model of Takeover Time

Kirti Mahajan
Research Scholar,
Transportation Systems Engineering,
Indian Institute of Technology, Mumbai, India,
Email: kirtimahajan.nith@gmail.com

David R. Large
Senior Research Fellow,
Human Factors Research Group,
University of Nottingham, Nottingham, UK,
Email: david.r.large@nottingham.ac.uk

Gary Burnett
Professor, Human Factors Research Group,
University of Nottingham, Nottingham, UK,
Email: gary.burnett@nottingham.ac.uk

Nagendra R. Velaga
Associate Professor,
Transportation Systems Engineering,
Indian Institute of Technology, Mumbai, India.
Email: velaga@civil.iitb.ac.in (*Corresponding author)

Word Count: 6984+ 2 tables (250 words per table) = 7484 words

(Submission Date: 24 July, 2020)

Paper Submitted for Presentation at the Transportation Research Board’s 100th Annual Meeting, Washington, D.C., 2021

ABSTRACT

The current study investigated the role of an in-vehicle digital voice-assistant (VA) in conditionally automated vehicles, offering discourse relating specifically to contextual factors, such as the traffic situation and road environment. The study involved twenty-four participants, each taking two drives: 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 signs, 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 31%. There was also some evidence that demographic factors influenced the TOT of drivers. Male drivers likely to resume control 1.72 times earlier than female drivers. The study findings highlight the benefits of adopting a futuristic in-car voice assistant to keep the drivers alert and aware about the recent traffic environment in partially AVs.

Keywords

Voice-user interfaces (VUI), conditional/partial automation, SAE Level 3, passive fatigue, driver takeover.

INTRODUCTION

The advances in vehicle automation allow the drivers to disengage from driving and become a passive monitor of the system. However, 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 at such critical moments (*1–4*). Such a decline in alertness, caused by 'task-disengagement' or 'low workload' conditions during automated driving, induces a passive fatigue state, resulting in loss of awareness of the current traffic situation(*1*, *5*, *6*).

Passive fatigue is specifically a problem in partially autonomous vehicles (level 2 or level 3 automation according to SAE International, (*7*)), in which the system will issue a takeover request (TOR) to the driver in situations that fall outside its capability. Ensuring a safe takeover of vehicle control is therefore one of the major challenges for highly automated driving. Hence, 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 (*8*).

Automation and alertness

Automation in vehicles can lead to long periods of driver inactivity leading to loss of alertness. Some studies showed a significant increase in symptoms of fatigue after only 15-20 minutes of automated driving in an *autonomous vehicle* (AV) (*2*, *3*, *8*). A study by Neubauer *et al.* (*2*) found that 30 minutes of automated driving increased proneness to fatigue, indicated by the mean Driver Stress Inventory (DSI). Gonçalves *et al.* (*10*) noticed that after only 15 minutes of automated driving there was significant increase in Stanford Sleepiness Scale (SSS) rating indicating increased sleepiness among drivers, leading to poor driving performance. Further, Neubauer *et al.* (*2*) found significant correlations between the fatigue ratings and lower task engagement using the Dundee Stress State Questionnaire (DSSQ). Saxby et al. (*6*) observed that the automation led to lower workload ratings on NASA-TLX scale(*11*) indicating low-workload. Wu et al. (*8*) reported a significant increase in eye blink duration and subjective sleepiness captured through Karolinska Sleepiness Scale (KSS) after 30-minutes of automated driving indicating fatiguing effects, irrespective of drivers' age. Vogelpohl *et al.* (*3*) noticed that 20 minutes of automated driving led to same level of fatigue among drivers as experienced after 40-50 minutes of manual driving, where fatigue was indicated using KSS sleepiness rating and indicators such as yawning, blinking, eye-closure etc. Collectively, these studies suggest that a period of automated driving lasting 20 minutes or longer is sufficient to significantly lower the task engagement and cause a significant decline in driver alertness.

Takeover time

The need for human intervention in partially automated vehicles is conveyed through pre-defined alerts, called as a takeover request (TOR). 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)* (*12*), and demonstrate their readiness to drive by re-engaging with the driving controls (*13*, *14*). Many studies have used the driver's response time to a hazard scenario at TOR as takeover time (*8*, *13*, *15–17*).

However, in the case of a potential collision event, the response is often reflexive, typified by sudden, emergency braking (*14*, *17*). Therefore, it would seem prudent to avoid a reflexive braking response to measure the absolute time to motor readiness when a TOR is issued, and therefore avoid a hazard scenario. This approach was taken by Merat et al. (*18*), 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.

Gaze behavior and situtation awareness

Passive fatigue due to mental underload situations can reduce the visual attention of drivers resulting in additional time required to regain self-alertness before building their SA in response to TOR (*3*, *17*). Visual, or gaze behavior can be a useful method to explore the process of allocating attention (*19*). Factors such as the duration and frequency of glances spent checking the speedometer, the road ahead, and side and rear-view mirrors, which are all associated with building situation awareness(*12*, *13*, *16*). Furthermore, a combination of hands on steering, placing feet on the pedals and looking at road ahead are indicative of motor readiness or readiness to drive (*13*, *14*, *17*). Few previous studies report driver's involvement in checking rear and side mirrors to gain situational awareness in response to a TOR (using a collision event at TOR)(*14*, *18*, *20*). Vogelpohl *et al.* (*3*) found that fatigue due to automation delayed drivers' reponse (i.e. resuming manual control) to a TOR by 5 to 8 seconds. Moreover, the first glance to speedometer (in the dashboard) was reported at 14 to 15 seconds after the TOR. The time required for visual and cognitive processing of the situation is therefore influenced by the driver's state of alertness (*17*). This highlights the need to develop driver assistance systems to keep the drivers engaged with driving and the traffic environment to ensure a safe and timely takeover.

Factors influencing takeover time

There are several demographic factors influencing takeover time after highly automated driving (*13*, *17*, *18*, *21*, *22*). For example, Warshawsky-livne and Shinar (*23*) found that gender made no significant difference on perception time, although male drivers had slower brake-movement time compared to females in response to brake lights in a leading vehicle. However, Mehmood and Easa (*24*) found that female drivers had longer reaction time than male drivers. Gomez et al. (*25*) found that women tended to exhibit more exploratory visual behavior, including longer fixations and scan path lengths during a task, and this could potentially lengthen takeover time. Nevertheless, Zeeb et al. (*17*) 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.

Most of the studies have used analysis of variance (ANOVA) technique to study the effect of these covariates on TOT (*8*, *13*, *15–17*). Zeeb et al. (*17*) adopted an integrated model approach to emphasize that primarily cognitive processes such as gaze behaviour determine the TOT. However, the study findings were limited to an emergency scenario post TOR without accounting for other factors e.g. gender, mileage etc. Therefore, it would be beneficial to model the takeover time, to quantify the contribution of such factors (age, gender, gaze behavior etc.) on the time to motor readiness at TOR. Existing studies in transportation research have widely used a parametric duration modelling approach to model the response time of drivers considering several covariates (*26*). As takeover time is a form of response time to the prompt of takeover request, this modelling approach can be extended to study the influence of factors such as alertness, gaze behavior etc. on takeover time.

Conversation to counter passive fatigue

Various studies have claimed the alerting effects of engaging in an active conversation for the drivers while driving (*27*, *28*). These benefits of conversation can be extended to mitigate the effects of passive fatigue during automated driving, and specifically the impact on the process of gaining motor readiness (*29*), (*30*). A few studies have proposed the use of a digital voice assistant(VA) conversing with the driver throughout the journey (*31*, *32*). These studies found that short intermittent conversations with a VA improved driver alertness and avoided potential driving performance decrements due to low alertness (*27*). Large et al. (*31*) showed that general 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. Thus, literature suggests that the content of conversation and frequency of exchanges is likely to play a significant role in avoiding fatigue without being distracting(*29*), (*30*). Drews, Pasupathi and Strayer, (*33*) 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, a speech-based interface or VA, could play the role of a driving coach or assistant, providing feedback to the driver about recent or upcoming traffic situations during automated driving. Such information may help the driver to effectively regain situation awareness during a takeover(*12*, *18*, *33*). However, providing traffic narrative at the point of takeover is likely to distract the driver and could influence their driving performance. An alternative would be to provide additional traffic information intermittently throughout a long journey. This is likely to improve driver alertness and awareness may additionally reduce the takeover time.

Study motivation and hypothesis

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 through traffic-related information. Therefore, 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. Secondly, 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, previous studies show the possible influence of various factors on TOT (age, experience, gender, involvement in secondary tasks, etc.) in isolation. However, they act as covariates to influence the takeover process. Further, the effectiveness of a VA proposed here, also depends on the ease and interest of drivers to use such technology. Therefore, this study focuses on modelling the TOT to determine how the presence of VA, (either directly or indirectly by influencing the gaze behavior or other factors gender or accustomed to use of voice assistants), can be effective in assisting takeover or resuming manual control after automated driving.

METHOD

The study was conducted using a medium 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. Three inobtrusive cameras were installed at different positions inside the car to capture drivers' hand and feet movements in response to TOR, and any physical signs of sleepiness during the experiments (yawning, extended blinks etc.). This simulator has been used in various previous studies (*27*, *31*), and is capable of providing an experience of driving level 3 automated vehicle. VA mainly conveyed driving-related information to the drivers, and thus, aimed to enhance the process of (re)gaining situation awareness, which was subsequently assessed using gaze behavior i.e. checking mirrors, road in front or traffic signs/signals etc. – factors that reflect the takeover time(TOT) among drivers (*14*, *17*). The TOT was modelled using a parametric duration approach illustrating the influence of VA and related situation awareness (as covariates) on TOT.

Participants

Participants were invited using flyers displayed around the University campus. A pre-driving questionnaire was used to identify and exclude participants reporting any sleep-related disorders and/or an Epworth Sleepiness Score (ESS) >16 (*34*). In addition, participants were specifically instructed to refrain from consuming caffeine, mint or alcohol for a few hours before the study and to take adequate sleep prior to the day of study. Thirty-one eligible participants volunteered for the study. Three participants did not turn up for the experiment and four dropped out partway through due to simulator sickness. The final data is reported from the remaining twenty-four participants (Table 1). Each participant received due compensation for their time. The study protocol was approved by the Faculty of Engineering Research Ethics Committee, University of Nottingham, UK.

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 the same urban scene. The posted speed limit varied from between 30-40mph in the urban scene to 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 environment such as buildings, types of cars etc. The roadside environment was intentionally sparse, with limited additional traffic, to minimise auditory and visual stimuli and increase monotony(*3*). Each drive lasted approximately 30minutes in the driving simulator with approximately 25 minutes of automated drive – considered to be sufficient to induce passive fatigue (*2*, *3*, *8*, *10*).

Figure 1 Experimental set-up: a. Driving Simulator and b. design view of dual carriageway

All participants drove in both the driving conditions (with and without VA) in a counterbalanced order to avoid any learning effects. A general layout of the experiment design and protocol is illustrated in Figure 2.

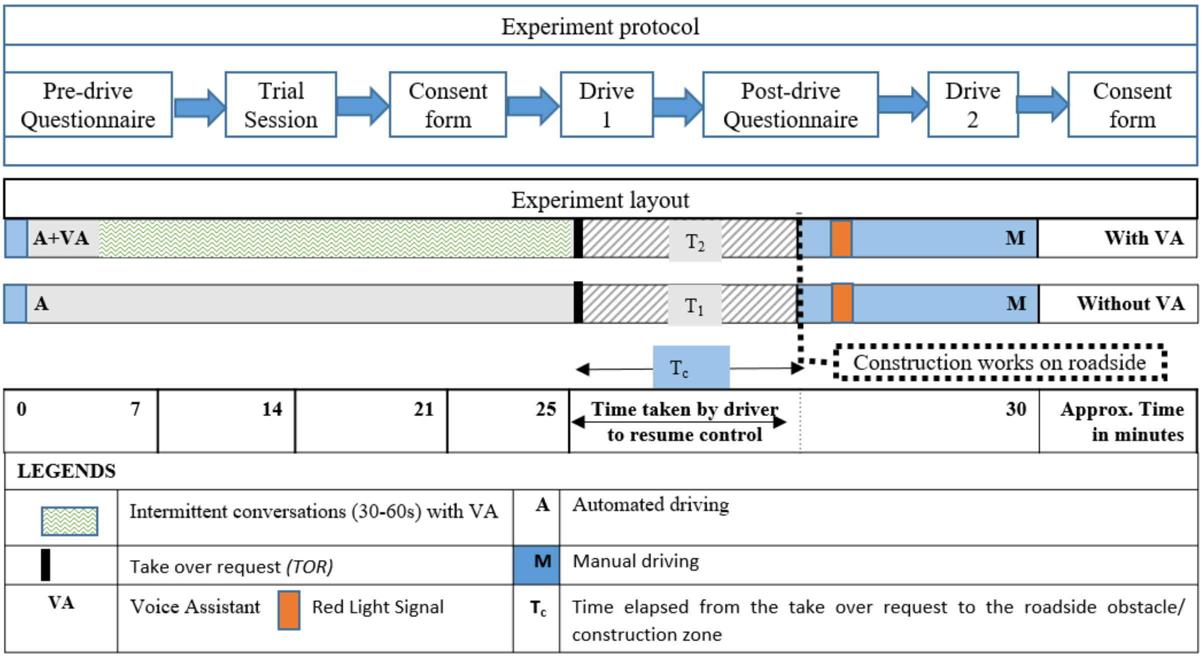


Figure 2 General layout of the experiment involving two drives on the driving simulator.

Drive with Voice Assistant (VA)

In the drive with VA (employing vocal interactions), VA introduced its capabilities such as providing surrounding traffic feedback, route navigation, event reminders or operating music or radio to the driver prior to the start of the drive. Drivers could either respond to or initiate conversation with VA using natural, conversational language. To achieve this, a comprehensive set of pre-recorded spoken messages were embedded in the STISIM scenario. These were played by the experimenter as per evolving scenario conditions, for example, changing speed limits, gap from leading vehicle, suggestions for rest/ refreshment spots, expected traffic congestion etc. The first message 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 (*8*). Each participant received the same opening gambits, however, the follow up statements differed slightly based on individual's response. For example,

VA: *"There is a pedestrian crosswalk ahead. Please be engaged in the drive or would you like to slow down?"*

Driver: "*yes*",

VA: *"Reducing speed. You are now driving at 'x' miles per hour."*,

However, this part of conversation will end if the driver responds *"no"*.

In situations where an appropriate reply was not available from the pre-recorded messages, VA responded with error messages such as "sorry! no network connectivity at the moment to perform this task", "sorry! This function is not available currently" (although in practice, these were rarely used). VA initiated a new conversational exchange or topic at approximately every 3 minutes, and these would last for at least 30-60 seconds (*29*). There was no conversation initiated during 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.

Takeover event

Prior to the test drives, participants undertook a practice drive involving multiple instances of takeovers, so that they become familiar with switching controls from 'manual' to 'automation', and vice-versa. The instruction to transfer control was intimated with a voice message followed by three consecutive beeps indicating the precise moment of the transfer of control, thereby avoiding any visual distraction. 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 (*35*). After completing 1-minute into the drive, automation was engaged at a fixed point at 0.75 miles (1.2km) in each test scenario. After approximately 25 minutes into the drive or at 21.6 miles (35 km) – participants received a takeover request (TOR). 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 at about 61m (200ft) from the onset of TOR with a gravel pile spilling on the roadside (Figure 3). In addition, a red-light traffic signal with a pedestrian cross-walk was presented at 152 m (approximately 500ft) from the point of TOR, and drivers were naturally expected to apply brakes in response to the signal. During the drive with VA, drivers received a notification few minutes prior to TOR, that they were approaching a pedestrian cross-walk.

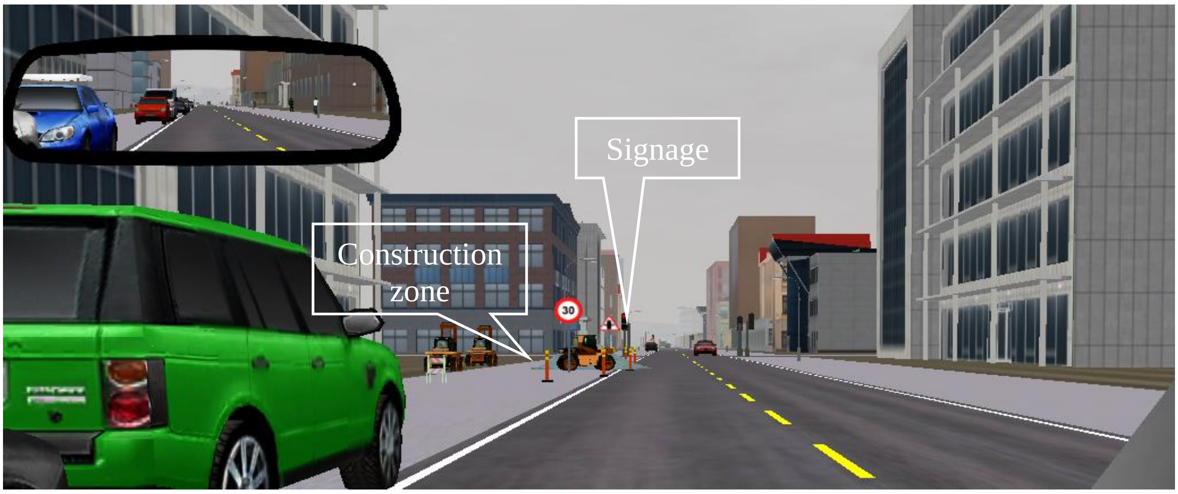


Figure 3 Scenario at takeover request (design view in STISIM at 100ft or 30.5m after onset of TOR)

Measures

Visual characteristics

SMI ‘natural gaze’ eye-tracking glasses were used to collect visual behavioral data at a sampling rate of 30 Hz. Pupil diameter, eye blink frequency and eye-blink durations were collected as indicators of fatigue during automation in each drive (*10*, *12*, *13*, *18*, *27*, *36*, *37*). 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). The following AOIs were selected: 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 (*12*, *13*, *18*, *36*). 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 (*17*).

Takeover time(TOT)

The TOT or response time to takeover request (TOR) 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, whereas the time to resume glances on road was noted from the eye tracking. The maximum of the three times was noted as the TOT.

Subjective sleepiness and workload ratings

In addition to the visual indicators of fatigue, drivers rated their level of alertness using the Karolinska Sleepiness Scale (KSS) (ranging from ‘1-very alert’ to ‘9-very sleepy’) and cognitive workload using NASA-TLX workload index (increasing scale of 1 to 21), (*2*, *11*, *27*, *28*). The subjective 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 during the drive, the latter two ratings were collected at the end of each drive. Also, the experimenter manually noted the relevant symptoms of sleepiness e.g. frequency of yawning and incidents of ‘nodding off’, when automation was engaged.

Post-drive Questionnaire

Finally, a questionnaire was used to collect data such as driver demographics, exposure to various in-car driver assistance systems (DAS) and voice-assistants, such as Google, Siri, Alexa etc. At the end of experiment, drivers were also asked to provide their subjective feedback on usefulness of VA.

ANALYSIS AND RESULTS

Dataset

The participant characteristics are summarized in Table 1. Most of the drivers already used voice-based assistants for either route navigation or as a music player (Figure 5), but did not use these to stimulate any conversations, for example, voice-based web search etc. In this study, individual responses to each suggested use (as listed in Figure 5) of a voice assistant (rated on a 1 to 5 Likert scale) were summed. This provided a single covariate indicating the frequency of using VAs rated on a linear scale varying from 1-20 (mean in Table 1).

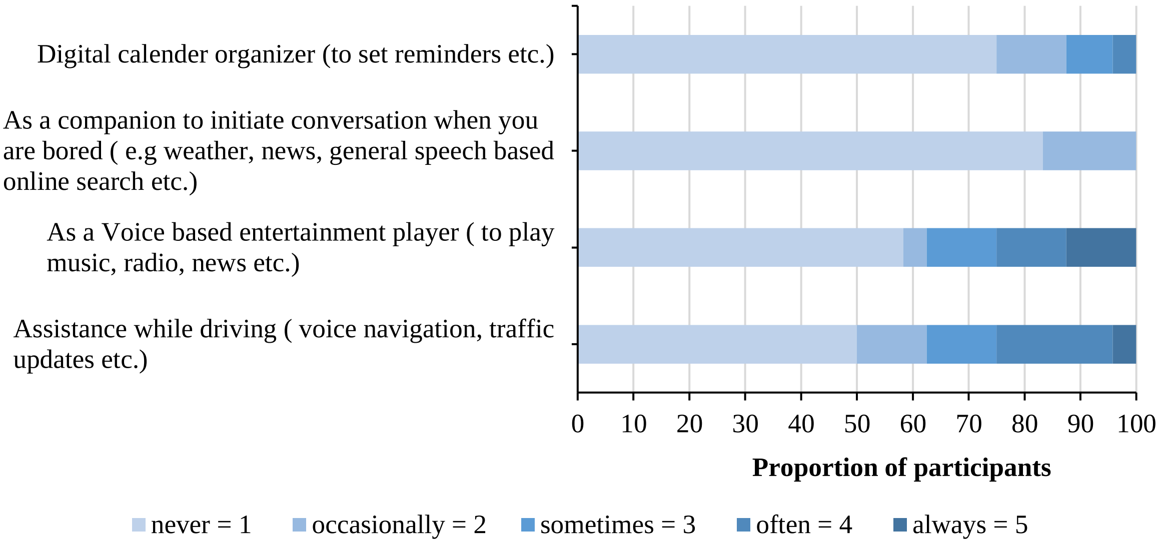


Figure 4 Frequency of using Voice Assistants such as Google, Alexa, Siri etc in different ways.

Table 1 also provides the paired t-test comparisons of subjective alertness, workload and response time to TOR in the two drives.

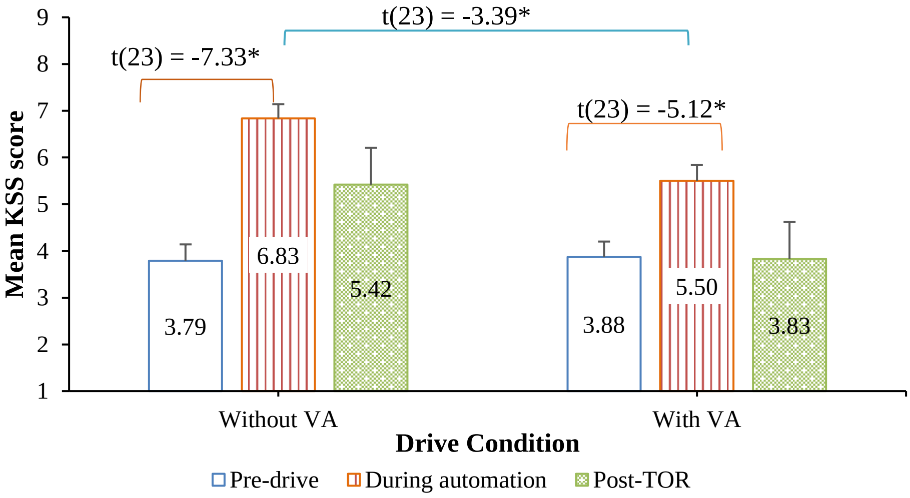
TABLE 1 Preliminary statistics of subjective data through questionnaires and visual indicators of sleep/fatigue

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			22	59	30.1	8.4
Annual mileage			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			5.00	9.00	7.2	0.9
Frequency of using VAs while in car (max.20)			4	14	6.8	3.2
Frequency of experiencing sleepiness symptoms while driving (max.40)			9	22	14.7	3.6
Subjective Rating (N=24)		**Min**	**Max**	**Mean**	**SD (±)**	**t(df)**
NASA TLX workload scale during automated drive	without VA	6	72	35.5	13.8	-2.45(23)*
	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	-2.26(21)*
	with VA	2.2	5.1	3.8	0.7	
Average eye blink frequency during automation (blink rate)	without VA	0.0	1.1	0.4	0.2	--
	with VA	0.0	1.2	0.5	0.2	
Response time to TOR (paired t-test comparison)				**Mean**	**SD (±)**	**t(df)**
Time to resume steering	without VA			3.2	1.57	2.01(22)**
	with VA			2.7	1.22	
Time to resume pedals	without VA			2.1	0.88	-1.95(21)**
	with VA			2.8	1.82	
Time of first glance at road ahead	without VA			2.9	1.46	--
	with VA			3.0	1.70	

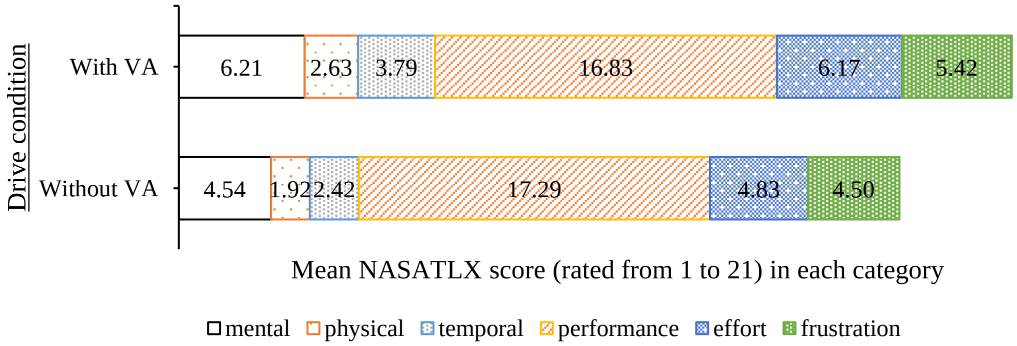
Note: t(df) indicate the pairwise t-test comparison between two drives at significance level '' (p<0.05), ** (p=0.05), '--' insignificant value (p>0.05)*

Alertness measures

The mean KSS and NASA-TLX workload ratings were compared using paired-samples t-tests (Figure 6 and Table 1). The t-tests showed a significant increase in mean KSS scores during automation (prior to TOR) from pre-drive in both the drives (Figure 6a). This indicates the fatiguing effects of automation, which sustain post-takeover as indicated by the KSS ratings, specifically in the absence of VA. However, the mean KSS rating during 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 NASA-TLX workload ratings were significantly higher during automation when the drivers were accompanied by VA 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. The paired comparisons of visual indicators of fatigue during automation i.e. average pupil diameter was significantly larger during the drive with VA, suggesting higher alertness with VA compared to the drive without VA.



a. Higher mean KSS scores during automation indicating loss of alertness (*indicating significant values i.e. p<0.005)

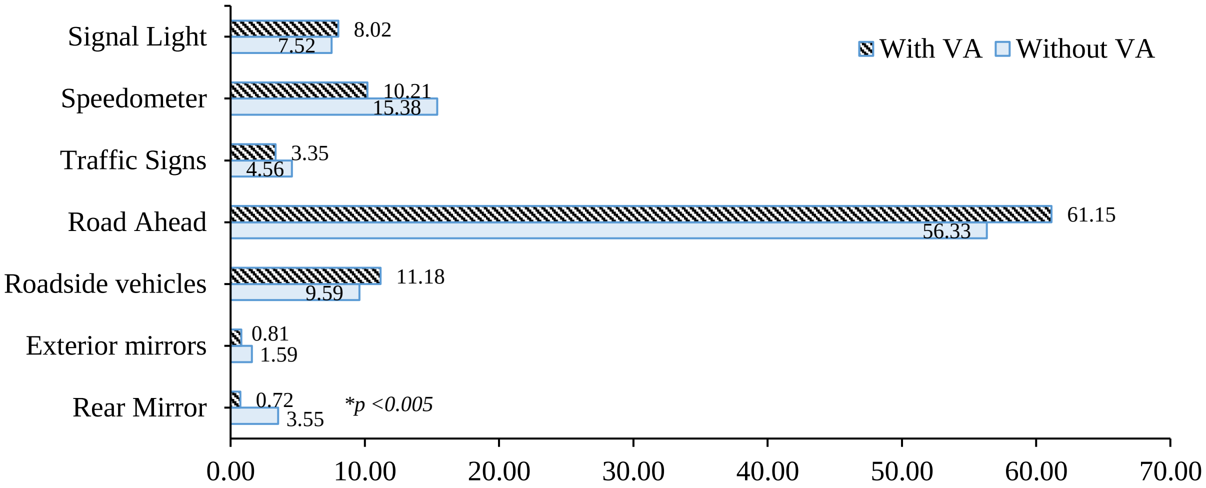


b. Increase in mean NASA workload ratings during automation in the drive with VA

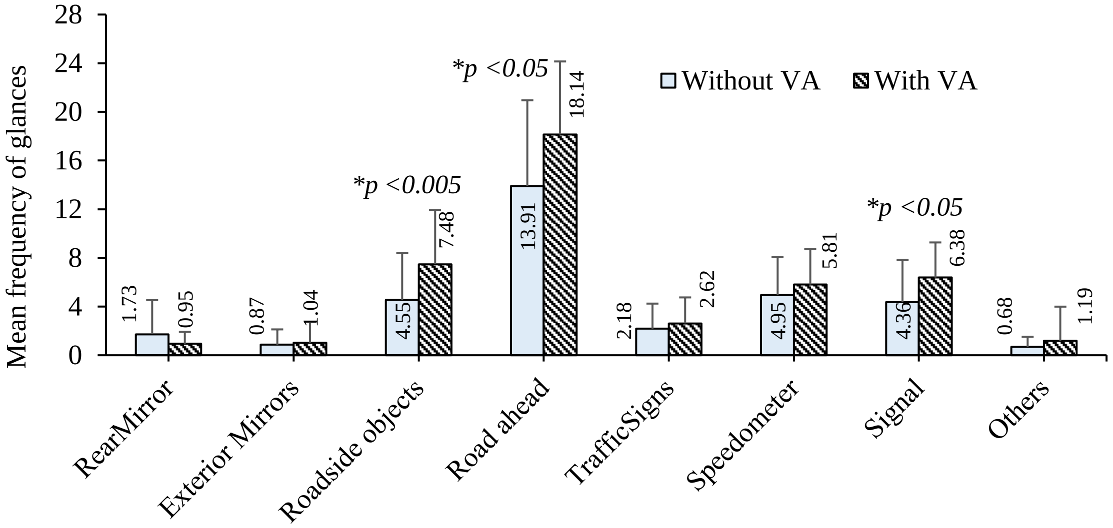
Figure 5 Low alertness and underload conditions observed during automation without VA

Situation awareness near TOR

The AOI data are compared in Figure 7 for the two drives. The percent time spent for 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 7. 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 traffic signage when provided in 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.



a. Average proportion of total glance duration



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)

Modelling takeover time (TOT)

The comparison of the alertness indicators and AOI statistics provide preliminary evidence of engaging effects of VA, which is likely to influence the takeover time. Therefore, a parametric duration model, or survival analysis approach was adopted to quantify the contribution of these factors on takeover time (*26*, *38*).

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 (*38*). 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).** The probability of resuming manual control after the time 't' (i.e. after the construction zone appeared) is called the **survivor function, S(t)**. 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 (*38*). In this study, accelerated failure time (AFT) model was used. An AFT model allows the covariates to rescale (accelerate) time directly in the baseline survivor function (*38*). Here, as the probability of completing the takeover is likely to increase over time, it indicates a monotone hazard rate that increases exponentially with time. Thus, Weibull distribution is suitable to model the takeover time data, with scale-parameter (P > 0) and location-parameter (λ > 0) is given by (*38*):

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

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

$$\mathrm{h}(\boldsymbol{t}) = (\lambda \mathrm{P})\ (\lambda \mathrm{t})^{P-1} \quad (2)$$

$$\mathrm{S}(\boldsymbol{t}) = EXP\ (-\lambda \mathrm{t}^{P}) \quad (3)$$

Here, the repeated observations were collected across the two drives with the same participants, which can cause intra-group heterogeneities. To account for such heterogeneities, Weibull AFT model with clustered heterogeneity and gamma frailty were developed and compared using Stata SE-16 (at 95% significance level). 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, with minimum Akaike's information criteria (AIC) and Bayesian information criteria (BIC) values (*38*) is reported in Table 2. The scale parameter p = 4.53 (>1) confirms that the hazard rate increased with time. Table 2 summarizes the estimated exponential of coefficients (hazard ratio) which directly represents the relative change in survival time duration with unit increment in the covariates. The model results show that participants were likely to gain motor readiness 31% quicker in the drive with VA compared to the other drive. There was a slight influence of cognitive workload, which was relatively higher during drive with VA. Also, the model results show that male drivers are likely to resume control 1.72 times earlier than female drivers. In addition, individuals who indicated that they frequently used VAs are likely to take 3% less time to resume control. Higher annual mileage and checking rear mirror did not influence the takeover time significantly.

TABLE 2 Weibull AFT (with gamma frailty) model estimates with the time to gain motor readiness following TOR as dependent variable

Variable	**Exp (B)**	**Coefficient (B)**	**Robust Std. Error**	**z**	**p-value**
1. Test condition					
With VA (compared to Without VA)	0.69	-0.374	0.156	-2.4	<0.05
2. Gender (female compared to male)	1.72	0.54	0.193	2.8	<0.001
3. Mileage	1.00	0.00	0.00	-1.69	<0.1
4. Using VAs	0.97	-0.033	0.017	-1.98	0.05
5. WL_TOR	0.99	-0.01	0.004	-2.43	<0.05
6. Likeliness of sleeping in AV	1.13	0.12	0.037	3.27	0.001
Glance frequency					
7. Road ahead	0.97	-0.032	0.022	-1.46	<0.1
8. Roadside	1.11	0.106	0.022	4.91	<0.001
9. Exterior mirrors	0.91	-0.099	0.039	-2.5	<0.05
%glance duration (or %time spent)					
10. Rear mirror	0.97	-0.028	0.029	-0.96	ns
11. Traffic signs	0.91	-0.095	0.019	-5.09	<0.001
Intercept		2.06	0.42	4.87	<0.001
Log psuedolikelihood		-4.61	-	p<0.001	
Scale parameter-p		4.53	1.35		
		N	df	AIC	BIC
Model estimates (with clustered heterogeneity)		24	13	35.21	50.52
Model estimates (with gamma frailty)		24	14	37.21	53.71

Note: ns: not-significant; VA: Voice assistant (Vid); WL_TOR: workload rating during automation (prior to takeover), AV: automated vehicle

The takeover probabilities were calculated for the two driving conditions (with and without VA) at different TOTs in Figure 8(a, b and c). All other variables were either kept at their reference category or corresponding means were substituted using Table 1, Figure 7.

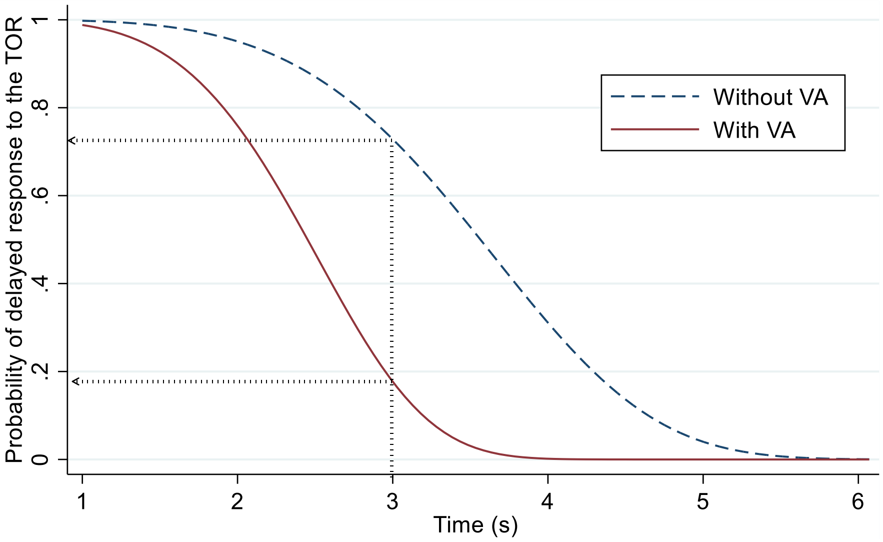


Figure 8a Survival curves for the two test conditions i.e. probability of not responding early to the TOR early

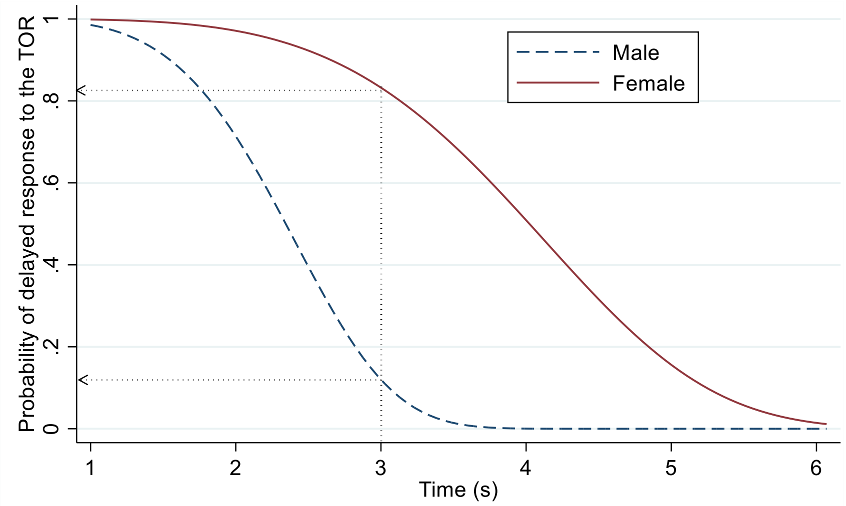


Figure 8b Effect of gender on probability of delayed response to takeover request

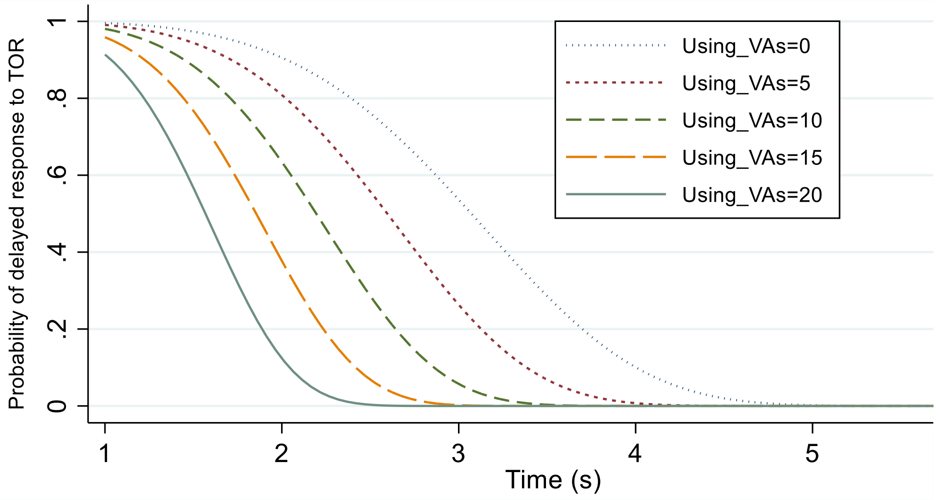


Figure 8c Reduction in probability of delayed takeover with higher frequency of using different types of voice assistants

DISCUSSION

VA and 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(*1*, *5*, *6*). Therefore, as hypothesized, the intermittent conversational exchanges with VA during automation led to a significant increase in driver workload as indicated by NASA-TLX ratings (Figure 6b). The lower KSS ratings and larger pupil dia in the drive with VA, indicated efficiency of VA in maintaining driver alertness during automation (*27*, *39*, *40*). Also, as noted by the experimenter, none of the drivers were observed sleeping during the drive with VA, whereas six drivers had short episodes of "nodding off" during automation in the drive without VA, and were notably startled by the takeover request. It is concluded that the regular conversational interludes made by the VA interrupted the monotony of the automated drive. 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.

Traffic feedback and situation awareness

The paired comparison of AOI statistics between the two drives 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 higher glances associated with checking exterior mirrors, concentrating on the road ahead and checking traffic signs during the drive with VA (Figure 7), relate to the verbal message delivered by VA (about the new speed limit and an approaching pedestrian crossing), a few minutes prior to TOR which shows that verbal cues can direct drivers' visual attention in the driving scene. Secondly, information about the pedestrians might have led to increased mirror-checks and additional focus made by drivers on the road ahead to prepare themselves for any remedial action that might be required. Increased mirror checks are generally associated with the cognitive processes to gain situation awareness at takeover (*14*, *17*, *21*). Drivers tended to shift their gaze to the speedometer, immediately after the posting of a new speed limit, resulting in no difference in glances at the speedometer between both driving conditions. However, it may also suggest a lack of trust and acceptance by drivers (for automation and the digital assistant), although this may change over time, as drivers' experience with such systems will increase. Another interesting finding was the increase in glances towards the rear-view mirror in the absence of VA. During the one-minute period of AOI analysis, there was no vehicle or event to reserve drivers' attention in the rear-view mirror which is otherwise a positive step in gaining SA (*17*, *36*). As mentioned by a few participants during a post-drive discussion, they were curiously observing the simulated objects in the scenario, indicating a potential distraction. Furthermore, without any vocal alerts to redirect their gaze at this time, drivers may have remained distracted by the virtual environment in rear-view mirror.

VA and takeover

For a timely takeover, the drivers were expected to check the surrounding environment prior to the construction zone to gain situation awareness (SA) and motor readiness by resuming driving controls, to avoid the risk of the car heading into the construction zone. The survival graph in Figure 8a shows the probability of a longer TOT is relatively higher during the drive without VA, compared to the drive with VA. During the drive with VA, drivers were more alert, to notice the construction zone following the TOR. Moreover, VA pre-informed them about an intersection signal ahead and the new speed limit, to engage them with driving environment, even in absence of any TOR. As shown by comparative AOI analysis (Figure 7), this information appears to have influenced drivers' gaze behavior, encouraging them to check for traffic signs, their speed etc., prior to TOR, thereby improving their ability to regain SA and reducing the TOT by 3% to 9% (see '*glance frequency*' in Table 2). Zeeb et al. (*17*) also claimed that gaze behavior is a significant indicator of cognitive process at TOR, influencing the TOT. However, during the drive without VA, drivers were not only fatigued and sleepy, but had been provided with no such traffic information. Therefore, it is suspected that the process of becoming alert and building SA would have been responsible for delaying the takeover process during this drive.

According to the model results, female drivers are likely to take longer to takeover (i.e. to demonstrate motor readiness) than male drivers (Figure 8b). 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 – similar results were reported by (*24*, *25*). 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 (*35*). Therefore, in this study, the drivers who expressed that they would be likely to sleep in an automated vehicle could suffer an increase in the probability of delayed takeover by 13% (Table 2). Nevertheless, willingness to sleep also suggests high trust and acceptance in the technology.

It was apparent that drivers who 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). The drivers expressed their intention to use similar voice-based driver assistant systems in the future, which is likely to have a positive impact on driver alertness and SA (*17*). A hypothetical increase in rating from 0 (drivers who have never used any voice assistant) to 20 (very often or always using different types of voice assistants) as shown in Figure 4, suggests an increase in adoption of such technology. The survival curves plotted in Figure 8c also suggest that higher use of these systems (indicated by increase in rating) could potentially increase their effectiveness in assisting the drivers during takeover after automation.

CONCLUSION

Extended periods of highly automated driving can disengage drivers from the driving task and reduce their alertness. Therefore, the AOI analysis and model findings show clear advantages of conversing with VA:

i. to counter the effects of passive fatigue.

ii. traffic-related information by VA can direct driver's cognitive process through relocating visual attention to traffic signs, mirrors or road-ahead.

iii. VA could effectively assist the drivers in a timely takeover.

Further, the parametric model of takeover time highlighted the gender-based differences in takeover time of drivers. The 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. However, the current study did not explore the effect of factors such as driver age, exposure to various in-car driver assistance systems due to limited sample size. The findings highlight the need of VA systems to maintain appropriate alertness and SA, especially for the drivers who may choose to sleep in highly automated vehicles. However, the positive effects of conversing with VA are likely to be transient, and therefore more research is required to investigate the lasting effects of such interventions.

ACKNOWLEDGEMENTS

This research was conducted through the financial support from Commonwealth Scholarship Commission (CSC) of UK under the split-site PhD research grant, 2018-2019. 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.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Gary Burnett, Kirti Mahajan, David Large; data collection and analysis: Kirti Mahajan. All authors reviewed the results, prepared the draft manuscript, and approved the final version of the manuscript.

REFERENCES

1. Matthews, G., C. Neubauer, D. J. Saxby, R. W. Wohleber, and J. Lin. Dangerous Intersections? A Review of Studies of Fatigue and Distraction in the Automated Vehicle. *Accident Analysis and Prevention*, No. April, 2018, pp. 0–1. https://doi.org/10.1016/j.aap.2018.04.004.

2. Neubauer, C., G. Matthews, L. Langheim, and D. Saxby. Fatigue and Voluntary Utilization of Automation in Simulated Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 54, No. 5, 2012, pp. 734–746. https://doi.org/10.1177/0018720811423261.

3. Vogelpohl, T., M. Kühn, T. Hummel, and M. Vollrath. Asleep at the Automated Wheel — Sleepiness and Fatigue during Highly Automated Driving. *Accident Analysis and Prevention*, No. July 2017, 2018, pp. 0–1. https://doi.org/10.1016/j.aap.2018.03.013.

4. Sheridan, B. T. B., and R. Parasuraman. Human-Automation Interaction. In *Reviews of Human Factors and Ergonomics*, pp. 89–129.

5. Di Milia, L., M. H. Smolensky, G. Costa, H. D. Howarth, M. M. Ohayon, and P. Philip. Demographic Factors, Fatigue, and Driving Accidents: An Examination of the Published Literature. *Accident Analysis and Prevention*, Vol. 43, No. 2, 2011, pp. 516–532. https://doi.org/10.1016/j.aap.2009.12.018.

6. Saxby, D. J., G. Matthews, J. S. Warm, E. M. Hitchcock, and C. Neubauer. Active and Passive Fatigue in Simulated Driving: Discriminating Styles of Workload Regulation and Their Safety Impacts. *Journal of Experimental Psychology: Applied*, Vol. 19, No. 4, 2013, pp. 287–300. https://doi.org/10.1037/a0034386.

7. SAE International. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. 2018.

8. Wu, Y., K. Kihara, Y. Takeda, T. Sato, M. Akamatsu, and S. Kitazaki. Effects of Scheduled Manual Driving on Drowsiness and Response to Take over Request: A Simulator Study towards Understanding Drivers in Automated Driving. *Accident Analysis and Prevention*, Vol. 124, No. January, 2019, pp. 202–209. https://doi.org/10.1016/j.aap.2019.01.013.

9. Matthews, G. Towards a Transactional Ergonomics for Driver Stress and Fatigue. *Theoretical Issues in Ergonomics Science*, Vol. 3, No. 2, 2002, pp. 195–211. https://doi.org/10.1080/14639220210124120.

10. Gonçalves, J., R. Happee, and K. Bengler. Drowsiness in Conditional Automation: Proneness, Diagnosis and Driving Performance Effects. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Vol. 31, No. 15, 2016, pp. 873–878. https://doi.org/10.1109/ITSC.2016.7795658.

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

12. Vlakveld, W., N. van Nes, J. de Bruin, L. Vissers, and M. van der Kroft. Situation Awareness Increases When Drivers Have More Time to Take over the Wheel in a Level 3 Automated Car: A Simulator Study. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 58, 2018, pp. 917–929. https://doi.org/10.1016/j.trf.2018.07.025.

13. Gold, C., D. Damböck, L. Lorenz, and K. Bengler. Take over! How Long Does It Take to Get the Driver Back into the Loop? *Proceedings of the Human Factors and Ergonomics Society*, Vol. 57, No. 1, 2013, pp. 1938–1942. https://doi.org/10.1177/1541931213571433.

14. Gold, C., R. Happee, and K. Bengler. Modeling Take-over Performance in Level 3 Conditionally Automated Vehicles. *Accident Analysis and Prevention*, Vol. 116, No. April 2017, 2018, pp. 3–13. https://doi.org/10.1016/j.aap.2017.11.009.

15. Neubauer, C., G. Matthews, and D. Saxby. Fatigue in the Automated Vehicle. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 58, No. 1, 2014, pp. 2053–2057. https://doi.org/10.1177/1541931214581432.

16. Samuel, S., A. Borowsky, S. Zilberstein, and D. L. Fisher. Minimum Time to Situation Awareness in Scenarios Involving Transfer of Control from an Automated Driving Suite. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2602, No. 2602, 2016, pp. 115–120. https://doi.org/10.3141/2602-14.

17. Zeeb, K., A. Buchner, and M. Schrauf. What Determines the Take-over Time? An Integrated Model Approach of Driver Take-over after Automated Driving. *Accident Analysis and Prevention*, Vol. 78, 2015, pp. 212–221. https://doi.org/10.1016/j.aap.2015.02.023.

18. Merat, N., A. H. Jamson, F. C. H. Lai, M. Daly, and O. M. J. Carsten. Transition to Manual: Driver Behaviour When Resuming Control from a Highly Automated Vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 27, No. PB, 2014, pp. 274–282. https://doi.org/10.1016/j.trf.2014.09.005.

19. Young, M. S., and N. A. Stanton. Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Human Factors*, Vol. 44, No. 3, 2002, pp. 365–375. https://doi.org/10.1518/0018720024497709.

20. Neubauer, C., G. Matthews, and D. Saxby. Fatigue in the Automated Vehicle: Do Games and Conversation Distract or Energize the Driver? *Proceedings of the Human Factors and Ergonomics Society*, Vol. 2014-Janua, No. 2009, 2014, pp. 2053–2057. https://doi.org/10.1177/1541931214581432.

21. Zeeb, K., A. Buchner, and M. Schrauf. Is Take-over Time All That Matters? The Impact of Visual-Cognitive Load on Driver Take-over Quality after Conditionally Automated Driving. *Accident Analysis and Prevention*, Vol. 92, 2016, pp. 230–239. https://doi.org/10.1016/j.aap.2016.04.002.

22. Louw, T., G. Markkula, E. Boer, R. Madigan, O. Carsten, and N. Merat. Coming Back into the Loop: Drivers' Perceptual-Motor Performance in Critical Events after Automated Driving. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 9–18. https://doi.org/10.1016/j.aap.2017.08.011.

23. Warshawsky-livne, L., and D. Shinar. Effects of Uncertainty , Transmission Type , Driver Age and Gender on Brake Reaction and Movement Time. *Journal of Safety Research*, Vol. 33, 2002, pp. 117–128.

24. Mehmood, A., and S. M. Easa. Modeling Reaction Time in Car-Following Behaviour Based on Human Factors. *International Scholarly and Scientific Research & Innovation*, Vol. 3, No. 9, 2009, pp. 325–333.

25. Gomez, P., A. Von Gunten, and B. Danuser. Eye Gaze Behavior during Affective Picture Viewing : Effects of Motivational Significance , Gender , Age , and Repeated Exposure. *Biological Psychology*, Vol. 146, No. June, 2019, p. 107713. https://doi.org/10.1016/j.biopsycho.2019.06.001.

26. Mahajan, K., and N. R. Velaga. Effects of Partial Sleep Deprivation on Braking Response of Drivers in Hazard Scenarios. *Accident Analysis & Prevention*, Vol. 142, 2020, pp. 1–22. https://doi.org/10.1016/j.aap.2020.105545.

27. Large, D. R., G. Burnett, V. Antrobus, and L. Skrypchuk. Driven to Discussion: Engaging Drivers in Conversation with a Digital Assistant as a Countermeasure to Passive Task-Related Fatigue. *IET Intelligent Transport Systems*, Vol. 12, No. 6, 2018, pp. 420–426. https://doi.org/10.1049/iet-its.2017.0201.

28. Matthews, G., J. Szalma, A. R. Panganiban, C. Neubauer, and J. S. Warm. *Profiling Task Stress with the Dundee State Questionnaire*. 2013.

29. Saxby, D. J., G. Matthews, and C. Neubauer. The Relationship between Cell Phone Use and Management of Driver Fatigue: It’s Complicated. *Journal of Safety Research*, Vol. 61, 2017, pp. 129–140. https://doi.org/10.1016/j.jsr.2017.02.016.

30. Atchley, P., M. Chan, and S. Gregersen. A Strategically Timed Verbal Task Improves Performance and Neurophysiological Alertness During Fatiguing Drives. *Human factors*, Vol. 56, No. 3, 2013, pp. 453–462. https://doi.org/10.1177/0018720813500305.

31. Large, D. R., G. Burnett, B. Anyasodo, and L. Skrypchuk. Assessing Cognitive Demand during Natural Language Interactions with a Digital Driving Assistant. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - Automotive’UI 16*, 2016, pp. 67–74. https://doi.org/10.1145/3003715.3005408.

32. Large, D. R., G. Burnett, D. Salanitri, A. Lawson, and E. Box. A Longitudinal Simulator Study to Explore Drivers’ Behaviour in Level 3 Automated Vehicles. *Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*, 2019, pp. 222–232. https://doi.org/10.1145/3342197.3344519.

33. Drews, F. A., M. Pasupathi, and D. L. Strayer. Passenger and Cell Phone Conversations in Simulated Driving. *Journal of Experimental Psychology: Applied*, Vol. 14, No. 4, 2008, pp. 392–400. https://doi.org/10.1037/a0013119.

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

35. Wan, J., and C. Wu. The Effects of Lead Time of Take-Over Request and Nondriving Tasks on Taking-Over Control of Automated Vehicles. *IEEE Transactions on Human-Machine Systems*, Vol. PP, 2018, pp. 1–10. https://doi.org/10.1109/THMS.2018.2844251.

36. Lu, Z., X. Coster, and J. de Winter. How Much Time Do Drivers Need to Obtain

Situation Awareness? A Laboratory-Based Study of Automated Driving. *Applied Ergonomics*, Vol. 60, 2017, pp. 293–304. https://doi.org/10.1016/j.apergo.2016.12.003.

37. Jamson, A. H., N. Merat, O. M. J. Carsten, and F. C. H. Lai. Behavioural Changes in Drivers Experiencing Highly-Automated Vehicle Control in Varying Traffic Conditions. *Transportation Research Part C*, Vol. 30, 2013, pp. 116–125. https://doi.org/10.1016/j.trc.2013.02.008.

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

39. Wang, X., and C. Xu. Driver Drowsiness Detection Based on Non-Intrusive Metrics Considering Individual Specifics. *Accident Analysis and Prevention*, Vol. 95, 2016, pp. 350–357. https://doi.org/10.1016/j.aap.2015.09.002.

40. Kapitaniak, B., M. Walczak, M. Kosobudzki, Z. Jóźwiak, and A. Bortkiewicz. Application of Eye-Tracking in Drivers Testing: A Review of Research. *International Journal of Occupational Medicine and Environmental Health*, Vol. 28, No. 6, 2015, pp. 941–954. https://doi.org/10.13075/ijomeh.1896.00317.