Exploring the effectiveness of a digital Voice Assistant to maintain driver alertness in partially automated vehicles

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

[a] *Transportation Systems Engineering, Indian Institute of Technology, Mumbai, India.*
[b] *Human Factors Research Group, University of Nottingham, Nottingham, UK,*

*Corresponding author: Transportation Systems Engineering, Indian Institute of Technology, Mumbai, India. Email: velaga@civil.iitb.ac.in

Exploring the effectiveness of a digital Voice Assistant to maintain driver alertness in partially automated vehicles

ABSTRACT

Objective: Vehicle automation shifts the driver's role from active operator to passive observer at the potential cost of degrading their alertness. This study investigated the role of an in-vehicle voice-based assistant (VA; conversing about traffic/road environment) to counter the disengaging and fatiguing effects of automation.

Method: Twenty-four participants undertook two drives– with and without VA in a partially automated vehicle. Participants were subsequently categorized into high and low participation groups (based on their proportion of vocal exchanges with VA). The effectiveness of VA was assessed based on driver alertness measured using Karolinska Sleepiness Scale (KSS), eye-based sleepiness indicators and glance behavior, NASA-TLX workload rating and time to gain motor readiness in response to take-over request and performance rating made by the drivers.

Results: Paired samples *t*-tests comparison of alertness measures across the two drives were conducted. Lower KSS rating, larger pupil diameter, higher glances (rear-mirror, roadside vehicles and signals in the drive with VA) and higher feedback ratings of VA indicated the efficiency of VA in improving driver alertness during automation. However, there was no significant difference in alertness or glance behavior between the driver groups (high and low-PR), although the time to resume steering control was significantly lower in the higher engagement group.

Conclusion: The study successfully demonstrated the advantages of using a voice assistant (VA) to counter these effects of passive fatigue, for example, by reducing the time to gain motor-readiness following a TOR. The findings show that despite the low engagement in spoken conversation, active listening also positively influenced driver alertness and awareness during the drive in an automated vehicle.

KEYWORDS

Conversation participation, voice user interfaces, passive fatigue, time to take-over, visual behavior

INTRODUCTION

Automation in vehicles has led to a significant shift in the driver's role from an active operator to a passive monitor of the system (e.g., level 3 automated vehicles; SAE International, 2018). Neubauer et al. (2014), Vogelpohl et al. (2018), and Wu et al. (2019) suggested that 15-20 minutes of such inactivity in an *automated vehicle* (AV) can significantly lower driver task-engagement (indicated by NASA-TLX ratings) and alertness indicated by increased eye blink duration, small pupil-diameter, frequent yawning, nodding off, subjective sleepiness scale ratings, etc. required to resume control.

The take-over time (TOT) is the response time for driver's assessment of the traffic situation or regaining adequate *situation awareness (SA)* (Vlakveld et al. 2018) to resume driving when a take-over request (TOR) is prompted (Zeeb et al. 2015). TOT is measured as the time to gain readiness to drive i.e. hands on wheel, feet on pedals and eyes on the road (Zeeb et al. 2015; Gold et al. 2018). Responding to a take-over request during passive fatigue conditions in an AV may require additional time to regain the required self-alertness and SA, thereby extending TOT (Vogelpohl et al. 2018). Therefore, there is a need for a system to keep the drivers alert and engaged with the driving environment during automation to ensure a timely and safe take-over (Samuel et al. 2016). Based on the efficiency of voice systems to grab driver's attention (Wong et al. 2019), this study investigated the role of a voice assistant to keep the drivers alert during a monotonous automated drive and assist them during the take-over by providing traffic updates or information (Wu et al. 2019).

Conversation and passive fatigue

An active conversation which engages the driver has been reported as an effective countermeasure to fatigue while driving (Jellentrup et al. 2001; Neubauer et al. 2014; Saxby et al. 2017; Large et al. 2018). Jellentrup et al. (2001) found 5-min phone conversations to be effective in mitigating fatigue. Neubauer et al. (2014) found that 10-min calls effectively increased driver workload during automation to avoid fatigue. However, to effectively mitigate fatigue, such conversations shall be intermittent and timed during the peak periods of fatigue (Atchley et al. 2013). Saxby et al. (2017) found that a 30s phone conversation on 'sharing past crash experiences' was distressing and ineffective in mitigating fatigue due to automation. It is feasible that, contrary to their study, longer conversations on general topics might help mitigate passive fatigue. Large et al. (2018) demonstrated that a natural language interface (NLI), engaging the drivers in general conversations about calendar reminders, news, music, etc. was an effective countermeasure to fatigue over cell phone conversations. The distracting effects of conversation can be avoided by naturalistic conversations on traffic (Drews et al. 2008). Therefore, some studies suggest the need for a system that can provide real-time traffic feedback to the drivers to maintain their alertness and increase situation awareness during automation (Vlakveld et al. 2018; Naujoks et al. 2019).

Goals and Hypothesis

In the wake of fatigue issues associated with automation, an in-car voice assistant (VA) was designed to provide traffic and route-related information, as well as general verbal interactions similar to those offered by voice assistants such as Siri, Cortana etc. (Large et al. 2018). It was hypothesized that a naturalistic conversation with VA would increase the cognitive workload such that the participants who would actively engage in vocal interactions with the VA, would be less likely to observe sleepiness symptoms. Secondly, the information and interactions provided by VA would keep the driver alert and may reduce the time to gain motor readiness at TOR. Thus, this paper aims to determine VA's effectiveness in mitigating passive fatigue due to automation and whether such benefits are dependent on driver’s vocal engagement with the assisting system.

METHOD

The study included two drives (with, and without, the VA) for each participant.

Participants

Twenty-four participants (10 females and 14 males) with a valid UK driving license participated in the study. The average age of participants was 30.1 years, and their mean driving experience was 10.5 years. Participants were asked to take adequate rest during the night before the study and to avoid consuming caffeine, mints, or alcohol before attending. Each participant received a £20 amazon voucher on completion of the study.

Experimental Settings

The study was conducted using a fixed-base, medium-fidelity driving simulator (Figure A1, online supplement), capable of simulating level 3 automation. STISIM Drive 3 software was used to develop the driving scenario. The scenario was presented on a curved screen comprising a 270 degrees field of view using three HD projectors. Each participant took two drives – with and without the voice assistant (VA) in a counterbalanced order. Both drives involved the same driving scenario of 25 miles comprising urban, rural, and dual carriageway elements, with a monotonous surrounding environment. Each scenario lasted around 30 minutes in the driving simulator with 25 minutes of the automated drive – considered to be enough to induce passive fatigue (Neubauer et al. 2014; Vogelpohl et al. 2018; Wu et al. 2019).

Intervention with Voice Assistant (VA)

During one of the drives, the participants were accompanied by VA, which employed natural language interactions based on the drivers' voice commands. The VA could provide driving assistance and information to keep drivers alert and engaged in driving, such as weather or traffic updates, route navigation etc. (examples are presented in Table 1). Conversational exchanges were pre-recorded messages embedded as sound files in the STISIM scenario and played by the experimenter as desired during the drive. The same opening gambits were delivered for each participant, starting after 5 minutes into the drive. VA-initiated a new conversation by either posing a question/information at approximately every 3 minutes to avoid signs of fatigue (Wu et al. 2019). The majority of these conversations (Table 1) included a set of follow-up questions/verbal exchanges which depended on driver's response to VA, and lasted about 30-60 seconds, playing the music/radio of driver's choice. For driver queries out of the scope of pre-recorded messages, VA responded with an error message, such as 'no network connectivity', 'function is currently unavailable' etc. However, such error messages were rarely used during the drive. There was no conversation during the 60 seconds prior to the TOR, although VA had already informed drivers to take caution of an upcoming change in the posted speed-limit and the approaching pedestrian crosswalk.

Driving task

Drivers were pre-informed that they might be required to resume control in response to a TOR. The participants practiced the transfer of controls from 'manual' to 'automation' and vice-versa at multiple TORs in a trial drive. In each test-drive, automation was engaged at a fixed point after an initial drive of 1 minute (0.75 miles) and a TOR was prompted in the final urban scene, after approximately 25 minutes (22 miles) into the journey. The TOR was presented using a voice message followed by three consecutive beeps to indicate the precise moment of automation disengagement. The scenario at TOR is shown in Figure A1(online supplement).

Measures

Engagement in conversation

The level of engagement in conversation was measured in terms of a participation ratio (PR), the words spoken by the drivers relative to VA's total words. The engagement in conversation was categorized into three levels based on the high, low and middle percentiles of PR. Out of the 24 participants in the drive with VA, the low PR group with participation ratio below the 40th percentile included 9 subjects, and the high PR group with participation ratio above the 60th percentile included 10 subjects. The remaining 5 subjects (with participation ratios between the 40th and 60th percentile) constituted the middle group. The intention was to maintain a significant difference in the proportion of conversation engagement between the high and low conversation groups. Therefore, only the high and low groups were considered in the analysis due to proportionate samples in each group and avoid misleading interpretations due to the small sample of the middle group (Berry 2002).

Alertness and workload

The drivers were asked to rate their alertness level using the Karolinska Sleepiness Scale (KSS) and workload using the NASA-TLX questionnaire (Neubauer et al. 2014; Large et al. 2018). KSS and NASA-TLX workload ratings were collected pre-drive, for the period of automation (just before TOR) and post-drive. The latter two ratings were both collected at the end of each drive to avoid any interference during the drive. Visual indicators of fatigue such as pupil diameter, eye blink frequency and eye blink duration were also collected using SMI ‘natural gaze’ eye-tracking glasses with BeGaze 3.7 software (Large et al. 2018; Wu et al. 2019). The experimenter manually recorded relevant sleepiness symptoms such as frequency of yawning and "nodding off" by the drivers during automation. These were later confirmed from the video recordings of experiments. Further, eye tracking data was analyzed during the 60s of automation immediately prior to the TOR to determine the allocation of visual attention (glance duration and frequency) on different areas of interest (AOIs) e.g., traffic signs and signals, road-ahead, rear and side view mirrors etc. These data were compared between the two groups (high and low PR) and across two drives (i.e. with or without VA).

The take-over time (TOT)

TOT in response to the take-over request (TOR) was calculated as the time elapsed from the end of the TOR voice message (start of beeps) to the time of gaining readiness to drive using frame-by-frame analysis of experiment videos. The TOT was taken as the maximum time to demonstrate all three actions i.e. the participants' hands on the steering, feet on pedals, and eyes on the road (Zeeb et al. 2015; Gold et al. 2018; Vlakveld et al. 2018).

Post-drive Questionnaire

The post-drive questionnaire was used to collect driver demographic details and the driver-acceptance rating for VA on a semantic scale of 1 to 5 on aspects, namely usefulness, pleasant, good, nice, effective, likable, assisting, desirable, and raises alertness (Laan et al. 1997).

Analysis

A within-subject, repeated measures *t*-test approach was adopted to compare the subjective and visual alertness, and takeover time across the two drives (Table 2 and A1, online supplement). The independent sample *t*-tests were used to compare the alertness and takeover time across the two participation groups. A statistical comparison of allocation of visual alertness was also conducted between the two PR groups to understand the influence of VA in increasing situation awareness. However, as the multiple *t*-tests expose the possibility of inflated type-I errors, a revised multivariate analysis of variance (MANOVA) was conducted with Bonferroni adjusted α-level (0.05/number of dependent variables) to compare the two PR groups (Berry 2002). A subjective analysis of the post-study feedback of VA further explored its usefulness and limitations, to be considered in the futuristic design of in-vehicle systems.

RESULTS

VA and alertness

Paired-samples *t*-tests show a significant increase in KSS rating and cumulative NASA-TLX workload rating during automation in both the drives, as indicated in Figure 1. Table 2 indicates a significant difference in KSS rating, average pupil diameter, frequency of closing eyes and "nodding-off" during automation and their respective effect size. The eye-tracking data for a short period during automation was missing for two participants due to some technical malfunctioning, therefore the visual alertness measures pre-TOR were reported from the remaining 22 drivers. Also, participants reported significantly higher workload ratings with VA (Table 2) but there was no significant difference in driver alertness or workload rating based on their participation ratio (Table A1-online supplement). Paired *Wilcoxon* signed rank tests of AOI characteristics across the two drives revealed higher glances concentrated at rear-mirror (z=-1.96, p=0.05), roadside objects (z = -2.08, p<0.05), signal (z = -3.18, p<0.001) and road ahead (z=-2.16, p<0.05) during the drive with VA (Figure 2). However, there was no difference in glance behavior or AOI characteristics between the two conversation-engagement (PR) groups.

VA and time to take-over

The paired *t*-tests showed no significant effect of VA on the mean TOT. However, the paired *t*-tests comparison of each component of takeover time, showed significant difference in time to resume steering and pedals between the two drives (Table 2). Each variable data was checked for outliers, which were removed prior to conducting statistical comparisons. Further, independent samples *t*-tests between high and low PR groups (Table A1, online supplement) showed significant difference in the time to resume steering and a near-significant difference in time to resume pedals, but no difference in time to resume glances at road ahead. The *t*-test was followed by MANOVA with PR as independent variable. Based on significant variables from *t*-test, dependent variables used were: NASA-TLX workload rating prior to the TOR, time to resume steering, time to resume pedals, time to resume glances on the road and proportion of conversation with VA. There was a statistically significant difference between PR groups on the combined dependent variables, (F (5, 13) = 5.65, p=0.006; *Wilk's* λ = 0.32; partial eta-squared (η_p^2)=0.69). However, the only differences to reach statistical significance, using a Bonferroni adjusted α-level of 0.01, were time to resume steering (F (1, 17) = 7.29, p = 0.005, η_p^2= 0.33) and participation ratio in conversation with VA (F (1, 17) = 14.86, p = 0.001, η_p^2= 0.47). The mean time to resume steering was 1.33s shorter for the high PR group compared to the low PR group.

Post-study feedback

The mean ratings evaluating the performance of VA are represented separately for the low and high PR groups in Figure A2 (online supplement). Twenty participants found that the VA was helpful in keeping them alert and indicated that they felt it assisted them when taking-over manual driving at TOR, stating that it, "provid[ed] useful information and assistance", was "a good companion when driving alone", and "help[ed] them to stay focused on road". In addition, participants commented that they "fell asleep without VA", that "[VA] kept [them] alert and proactive", that they "like[d] when [VA] provided relevant traffic details", that "information like [the] speed limit was helpful in take-over", and that VA "ke[pt] [them] engaged with driving instead of checking phones" etc. Four participants who did not like the system, three from the low PR group, stated reasons such as "it was annoy[ing] with frequent interruptions", and that they "d[id] not like to interact with devices", although one commented that they "enjoyed only the bit with music". In addition, the participant from the high PR group stated that VA was "temporarily alerting only during conversation, but, then [made them] more drowsy by increasing dependence on VA to keep [them] engaged". There was no significant difference in the overall performance rating of VA among high and low participation groups (Table A1-online supplement). However, further inspection of mean ratings on each component of the scale in Figure A2(online supplement) shows that drivers in the high PR group (high verbal engagement with VA) found VA to be more efficient in raising alertness and as a driving assistant. Despite the low verbal engagement, participants from the low PR group reported VA to be more useful and a good device to be used.

DISCUSSION

The study demonstrated the advantages of using a voice assistant (VA) to counter the effects of passive fatigue. Lower reported KSS ratings and higher pupil diameter indicate higher alertness in the drive with VA supporting the first hypothesis (Körber et al. 2015; Large et al. 2018). In the drive with VA, drivers reported significantly higher cognitive workload due to interactive conversations which significantly reduced the sleepiness symptoms (Vogelpohl et al. 2018). None of the drivers were observed nodding-off during the drive with VA, whereas six drivers nodded-off during automation in the non-VA drive. However, contrary to our hypothesis, such improvement in alertness due to VA was irrespective of the extent of vocal participation. It is of course feasible that drivers could chose to listen to traffic–related information, such as speed limits, upcoming intersections/crossings, etc., without offering a verbal response, and this alone could contribute towards improving their awareness of the road situation.

For the second hypothesis, the comparisons across two drives showed no significant difference in the overall takeover time. This may be so, as the mean time to resume contact with the foot pedals increased as much as the time to resume steering reduced in the drive with VA. The VA was aimed at providing timely, relevant information to the drivers to keep them prepared for a safe and measured take-over of control. Thus, it is possible that drivers resumed the pedals at their ease on receiving the traffic information briefed by VA. Further, the average time to resume steering controls was 1.33s shorter in the high-PR group as compared to the low-PR group. This shows the possibility that the response to TOR can be significantly influenced by improved situational awareness due to traffic-related information such as an approaching intersection signal, etc. The significantly higher proportion of words spoken by drivers in the high-PR group suggests that drivers may have paid more attention to VA than the low-PR group, leading to a prompt response and reduced TOT. In the drive without VA, the drivers were feeling tired/sleepy. Consequently, the process of attaining sufficient alertness to gain the required level of situation awareness specifically among the low PR group delayed their resumption of steering, but conversely may have resulted in a sudden and instinctive brake response (Zeeb et al. 2015; Gold et al. 2018).

The post-study subjective evaluation of VA also supports the effectiveness of using such a system to improve alertness and situation awareness during automation. Figure A2(online supplement) shows that high PR group found VA to be more efficient as a driving assistant and effective in raising alertness compared to low PR group (Atchley et al. 2013). However, most of the drivers expressed an interest in the entertainment functions of VA which could be distracting even for drivers in low PR group, thereby delaying their responses (Choudhary & Velaga 2018). Another limitation pointed out by a low PR group driver was the frequent interruptions by VA, which might have annoyed drivers, making it unpleasant or less likable to use (Figure A2, online supplement). It is also worth mentioning that prior experience of using other voice assistants such as Google, Siri, etc. might also influence

VA's usability. The negative feedback suggests that drivers may get bored or annoyed by repeated conversations offered by such systems over time. Alternatively, the positive feedback indicates that they may appreciate such a system's benefits, increasing its usability and efficacy over time. Also, as pointed out by a participant, the alerting effects of conversation could be limited to the period of conversations. Therefore, to improve the acceptance and usability of such interfaces, it is suggested that they should be capable of adapting to driver's interests and expectations, in addition to providing traffic-related feedback. Nevertheless, all drivers acknowledged the usefulness of VA in keeping them alert during automation.

A significant limitation of current study is the limited sample size and neglecting the middle group. Future studies with higher sample size may use regression techniques to model the effect of conversation engagement (as a continuous variable), thereby including all samples. The proportion of vocal conversation was significantly higher in the high PR group than the low PR group, but future studies should also consider alternative methods to include "listening" as part of conversation-engagement. Moreover, other factors, such as age, may also influence driver alertness, the use of VA and take-over performance (Wu et al. 2020). Further, the effectiveness of providing traffic feedback in preparing the drivers for take-over may be investigated by tracking the corresponding visual behavior. The findings aim to demonstrate that providing traffic information or other conversations with a VA can heighten drivers' alertness and prepare them for a safe take-over of control.

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

Data Availability
The datasets used in the current study are protected by relevant general data protection regulation(GDPR), University of Nottingham policies. Selected, anonymised datasets may be available from the authors upon reasonable request.

REFERENCES

Atchley P, Chan M, Gregersen S. 2013. A Strategically Timed Verbal Task Improves Performance and Neurophysiological Alertness During Fatiguing Drives. Hum Factors. 56(3):453–462.

Berry, MJ. 2002. A Step-by-Step Guide to SPSS for Sport and Exercise Studies.

Choudhary P, Velaga NR. 2018. Performance Degradation During Sudden Hazardous Events: A Comparative Analysis of Use of a Phone and a Music Player During Driving. IEEE Trans Intell Transp Syst. PP:1–11.

Drews FA, Pasupathi M, Strayer DL. 2008. Passenger and Cell Phone Conversations in Simulated Driving. J Exp Psychol Appl. 14(4):392–400.

Gold C, Happee R, Bengler K. 2018. Modeling take-over performance in level 3 conditionally automated vehicles. Accid Anal Prev [Internet]. 116(April 2017):3–13. https://doi.org/10.1016/j.aap.2017.11.009

Jellentrup N, Metz B, Rothe S. 2001. Can talking on the phone keep the driver awake ? Results of a field- study using telephoning as a countermeasure against fatigue while. (1995):1–12.

Körber M, Cingel A, Zimmermann M, Bengler K. 2015. Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving. In: Procedia Manuf. Vol. 3; p. 2403–2409.

Laan JD Van der, Heino A, de Waard D. 1997. A Simple procedure for the assessment of acceptance of advanced transport telematics. Transp Res Part C Emerg Technol. 5(I):1–10.

Large DR, 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(6):420–426. http://digital-library.theiet.org/content/journals/10.1049/iet-its.2017.0201

Naujoks F, Hergeth S, Wiedemann K, Schömig N, Forster Y, Keinath A. 2019. Test procedure for evaluating the human–machine interface of vehicles with automated driving systems. Traffic Inj Prev. 20(sup1):S146–S151. https://doi.org/10.1080/15389588.2019.1603374

Neubauer C, Matthews G, Saxby D. 2014. Fatigue in the automated vehicle: Do games and conversation distract or energize the driver? Proc Hum Factors Ergon Soc. 2014-Janua(2009):2053–2057.

Samuel S, Borowsky A, Zilberstein S, Fisher DL. 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(2602):115–120. http://trrjournalonline.trb.org/doi/10.3141/2602-14

Saxby DJ, 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. http://dx.doi.org/10.1016/j.jsr.2017.02.016

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

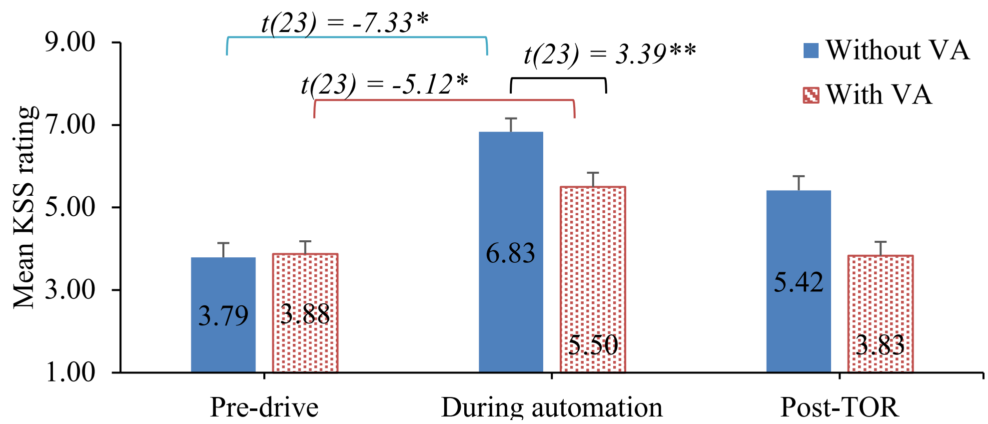
Vogelpohl T, Kühn M, Hummel T, Vollrath M. 2018. Asleep at the automated wheel — Sleepiness and fatigue during highly automated driving. Accid Anal Prev.(July 2017):0–1. http://dx.doi.org/10.1016/j.aap.2018.03.013

Wong PNY, Brumby DP, Ramesh Babu HV, Kobayashi K. 2019. Voices in self-driving cars should be assertive to more quickly grab a distracted driver's attention. Proc - 11th Int ACM Conf Automot User Interfaces Interact Veh Appl AutomotiveUI 2019.:165–176.

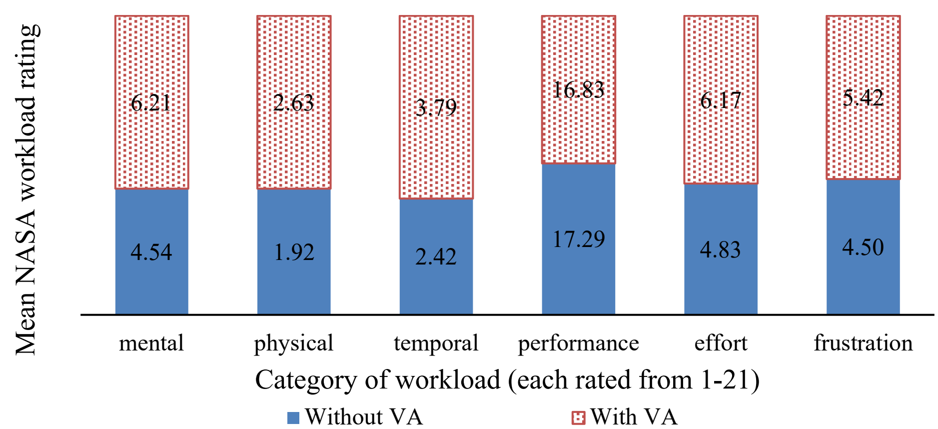
Wu Y, Kihara K, Hasegawa K, Takeda Y, Sato T, Akamatsu M. 2020. Age-related differences in effects of non-driving related tasks on takeover performance in automated driving. J Safety Res [Internet]. 72:231–238. https://doi.org/10.1016/j.jsr.2019.12.019

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(January):202–209. https://doi.org/10.1016/j.aap.2019.01.013

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. http://dx.doi.org/10.1016/j.aap.2015.02.023



a. Higher KSS Score indicating higher fatigue during automation in both drives (*p<0.001, **p<0.01)



b. VA increased the workload during automation (i.e. pre-TOR) as compared to workload during automation in drive without VA ($t(23)=-2.25$, $p<0.05$)

Figure 1 Mean KSS and NASA-TLX workload rating compared across two drives.

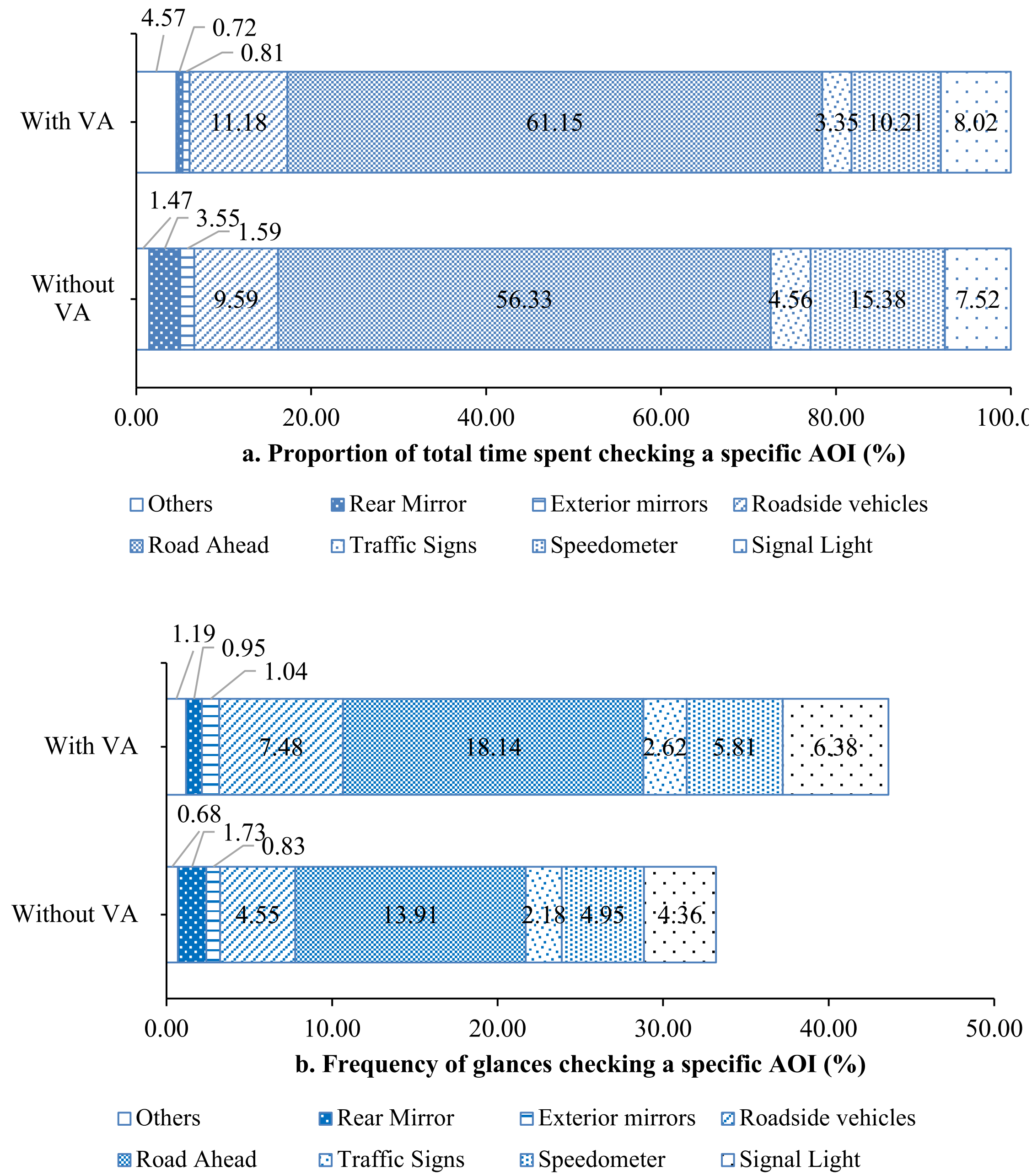


Figure 2 Representation of visual glances at defined areas of interest (AOIs) in the two drives during 60s prior to TOR.

Table 1 Examples of VA's opening gambits

Category	Example Statements
1. Event reminders (calendar)	"…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? …
3. Traffic/ Road	
i. Fatigue/Rest	"You have been driving for a long time today, would you like to stop for refreshments or rest? …"
ii. General journey	"We are currently headed towards 'A5250'. The posted speed limit is '60' miles per hour…"
iii. Road-traffic feedback	"There is a pedestrian crosswalk ahead. Please be engaged in the drive or would you like to slow down?"

Note: the entire conversation including follow-up statements was not limited to given example statements.

Table 2 Paired sample *t*-tests across two drives: with and without VA.

Variable	*Mean (SD)*		*t*	*df*	*p-value*	η_p^2
	Without VA	*With VA*				
1. KSS score at the end of automation prior to TOR	6.8 (1.63)	5.5 (1.72)	3.39	23	0.003	0.300
2. Average pupil diameter in mm during automation	3.6 (0.81)	3.8 (0.69)	-2.26	21	0.034	0.210
3. SS2: frequency of eye closing over 1s	1.9(2.38)	0.5 (1.18)	-2.45	23	0.022	0.255
4. SS3: frequency of nodding-off	0.6 (1.35)	0	-2.11	23	0.045	0.188
5. NASA TLX workload rating at the end of automation prior to TOR	35.5 (13.78)	41.0 (12.05)	-2.45	23	0.022	0.255
6. Time to gain motor readiness post-TOR	3.5 (1.53)	3.3 (1.51)	0.33	22	0.748	0.002
7. Time to resume steering	3.2 (1.57)	2.7(1.22)	3.01	22	0.050	0.289
8. Time to resume pedals	2.1 (0.88)	2.8 (1.82)	-1.95	21	0.064	0.176
9. Time of first glance at road ahead	2.9 (1.46)	3.0 (1.7)	-0.19	23	0.652	0.001

SD: Standard deviation;

Table A1. Independent sample *t*-tests between participation groups in drive with VA

Variable	Mean (SD) Low	Mean (SD) High	*t*	*df*	*p-value*	η_p^2
1. KSS score at the end of automation prior to TOR	5.00 (1.50)	6.00 (1.56)	-1.42	17.00	0.174	0.11
2. Average eye blink frequency during automation	0.44 (0.19)	0.5 (0.33)	-0.44	17.00	0.665	0.01
3. Average pupil diameter in mm during automation	3.83 (0.8)	3.64 (0.55)	0.61	17.00	0.551	0.02
4. NASA TLX workload rating at the end of automation prior to TOR	33.78 (12.18)	44.70 (8.72)	-2.27	17.00	0.037	0.29
5. Time to gain motor readiness post-TOR	3.96 (1.52)	2.92 (1.54)	1.48	17.00	0.157	0.11
6. Time to resume steering	3.39 (1.33)	2.06 (0.77)	2.70	17.00	0.007*	0.30
7. Time to resume pedals	3.38 (1.71)	2.11 (1.34)	1.83	17.00	0.050	0.20
8. Time of first glance at road ahead	3.33 (1.88)	2.72 (1.71)	0.74	17.00	0.469	0.03
9. Participation ratio in conversation with VA	8.69 (3.43)	30.95 (16.97)	-3.85	17.00	0.001*	0.48

SD: Standard deviation and * indicates the variables found significant after Bonferroni correction or adjusted alpha level MANOVA analysis.

Figure A3 Experimental set up: fixed-base simulator and scenario at take-over request (design view in STISIM at 100ft after onset of TOR)

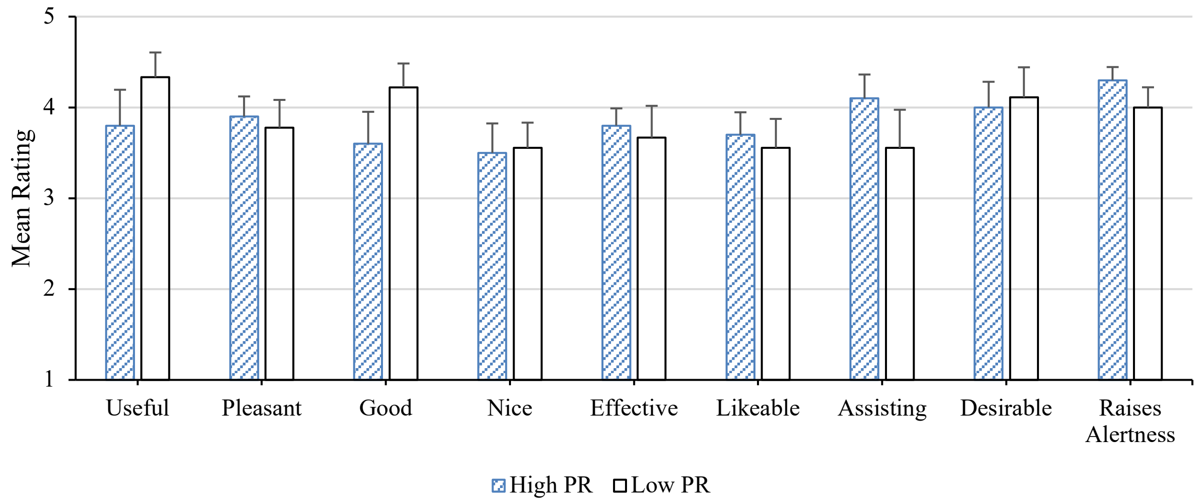


Figure A2. Post-study feedback for the performance evaluation of VA across the two participation ratio (PR) groups (vertical bars represent standard errors)