Highlights

Exploring the usage of supervised driving automation in naturalistic conditions

Jork Stapel, Riender Happee, Michiel Christoph, Nicole van Nes, Marieke Martens

- A naturalistic dataset comparing manual and SAE2 automated driving was enriched with automation status and driver attention
- Effects of road type, driving speed, time of day, trip duration and experience on automation use were examined
- Automation was used around 60% of time on highways and was rarely activated on roads with speed limits below 70km/h
- Automation use increased with time in trip while showing minor changes with time of day and automation experience
- Automation use and automation experience showed small but significant effects on head pose variance, indicating subtle changes in monitoring strategies

Exploring the usage of supervised driving automation in naturalistic conditions

Jork Stapel[a], Riender Happee[a], Michiel Christoph[b], Nicole van Nes[a,1], Marieke Martens[c,1]

[a] *TU Delft University of Technology, The Netherlands*
[b] *SWOV Institute for Road Safety Research, The Hague, The Netherlands*
[c] *TNO Traffic & Transport, The Hague, The Netherlands*
[d] *Eindhoven University of Technology, The Netherlands*

Abstract

This study reports usage of supervised automation and driver attention from a longitudinal naturalistic driving study. Automation inexperienced drivers were provided with instrumented vehicles with adaptive cruise control (ACC) and lane keeping (LK) features (SAE level 2). Data was collected comparing one month of driving without support to two months where drivers were instructed to use automation as desired.

On highways, automation was used respectively 63% and 57% of the time by Tesla and BMW users, increased with driving speed and was used the least while driving 10-60 km/h, where especially ACC usege reduced. On roads with speed limits below 70 km/h, automation was used less than 6%, and use on urban roads was incidental rather than habitual. Automation usage was higher during commute hours compared to other moments for the BMW group, and increased slightly with time in trip. Head pose without gaze data could not classify driver attention. Head pose deviation was selected as alternative indicator for monitoring activity. Comparing among forms of automation usage on the highway, head heading deviation was smallest during ACC use, but did not differ between automation and baseline manual driving. Head heading deviation during manual driving was smaller in the baseline than the experimental phase, which suggests that motives for manual highway driving may be attention related. Automation usage did not change much over the first 12 weeks of the experimental condition, and there were no longitudinal changes in head pose deviation. We recommend not to rely on head-pose for attention classification and consider gaze instead.

Keywords: automation use, driver attention
PACS: 0000, 1111
2000 MSC: 0000, 1111

1. Introduction

Supervised, or SAE Level 2 partial automation (SAE, 2021) is rapidly deployed in commercial cars. Current systems automate longitudinal control with adaptive cruise control (ACC) and support lateral control with lane keeping (LK). While Level 2 automation is active, the driver has to supervise the automation, and intervene when needed to ensure safety. Safe use requires that the driver is aware of these responsibilities, has an accurate understanding of how the vehicle may respond to the situation at hand, and maintains sufficient situation awareness to respond when necessary. However recent accidents indicate that drivers are not always monitoring the environment sufficiently (Dutch Safety Board, 2019). Harms et al. (2020) found that drivers are not always aware of the abilities and limitations of current systems. There are also clear indicators that automation can have a negative effect on driver vigilance, for instance when sleep-deprived, though automation can initially improve alertness when well-rested (Ahlström et al., 2021). If and how drivers experience system limitations can be inferred from how these systems are used, and how drivers monitor the automation distributing visual attention between driving-related and other tasks. Several studies have used such measures to evaluate safe usage of automated driving features. Jamson et al. (2013) found that the use of driving automation reduced the number of lane changes, whereas drivers spent more time on secondary tasks but adjusted their attention to the road depending on traffic. Similarly, Naujoks et al. (2016) demonstrated in a 2013 Mercedes E-class that drivers with prior ACC experience perform more secondary tasks while using driving automation, whereas drivers without ACC experience did not. Farah et al. (2021) found that drivers over-estimated the operational design domain as defined by the vehicle manufacturer in an on-road study with a Tesla model S. Banks et al. (2018) performed thematic video analysis of behaviours observed during on-road driving in a Tesla model S and identified multiple occurrences of missed notifications from the HMI leading to mode confusion.

The distribution of visual attention between driving-related and secondary tasks can be inferred from gaze or head movement (Lee et al., 2018) and provides guidelines for driver distraction from in-vehicle displays (Strickland, 2013). Park et al. (2017) demonstrated that a reduction of on-road glance duration impairs hazard detection performance. Glaser et al. (2017) demonstrated that eyes-off-road time negatively impacts driver performance when resuming manual control in critical scenarios. Additionally, gaze can be indicative of cognitive load or distraction (Wang et al., 2014), fatigue and intoxication (Victor et al., 2005).

As drivers' understanding of the automation develops with experience, so will their usage and monitoring behaviour (Sullivan et al., 2016). On road and simulator studies demonstrated substantial differences between drivers with and without driving automation experience. Larsson et al. (2014) compared control transition performance in a simulator between ACC users and drivers without prior driving automation experience and found that experienced users responded faster in cut-in scenarios. Victor et al. (2018) however demonstrated in a 30 minute test track drive in a Volvo XC90 that expectation mismatch during first-failures can result in a crash even with attentive drivers. Stapel et al. (2019) demonstrated that Tesla owners experienced with driving automation perceive a lower workload during automation use compared to first-time users. However, this perception was contrasted by a slower response time on an auditory detection-response task, indicating an increased objective workload when using automation. Hancock and Matthews (2018) provide further reflections on the occurrence of this dissociation. Large et al. (2019) performed a 5-day longitudinal simulator study on conditional automation without supervision showing high automation usage, trust and secondary task uptake. Time spent attending the road during automation use reduced from 30% to 20% over the five days accompanied with reduced driving performance after resuming manual control. The latter improved after introducing a routine for regaining situation awareness.

While several studies were conducted in controlled or semi-controlled on-road conditions, only few investigated the use of and adaptation to automated driving in a naturalistic setting. Beggiato et al. (2015) performed a longitudinal on-road study where they found that drivers developed their trust and functional understanding of ACC over ten drives while establishing a high acceptance within two drives. Morando et al. (2019) investigated how SAE2 driving automation influences attention during 10 months of naturalistic manual and automated driving by 17 participants. They found longer on-road glances and lower percent eyes on road centre during automated driving (ACC and LK) compared to manual driving. The latter was interpreted as a reduced task demand during automation use. Russel et al. (2018) conducted a naturalistic driving study with 120 participants driving vehicles equipped with adaptive cruise control and automated lane keeping for 4 weeks. They report effects of traffic stability, road type and weather conditions (no-precipitation vs precipitation) on automation use and found that drivers were performing secondary tasks 60% of the observed time regardless of automation use and found no difference in percentage eyes-off-road time, off-road glance duration or type of secondary task. Reaction times to the 'hold steering wheel'- requests did not change over the four weeks of use, but instances occurred in the first week where such requests were intentionally ignored to investigate the vehicle's response. While these studies provide useful insights, the evolution of behaviour from manual to automated driving has mainly been examined for the first experience with automation, or lack observations of baseline manual driving prior to developing experience with automated driving.

In this study we report automation use and driver attention from a longitudinal naturalistic driving study conducted in the Netherlands. The study is unique in its inclusion of a one month manual driving baseline followed by a two month experimental phase with the same participants and vehicles where participants were allowed to use the vehicle's automation, enabling a within-subject analysis of behavioural adaptation over the first two months of automation usage.

We addressed the following three research questions:

1. When and in which conditions do drivers use ACC and LK support?
2. Is driver attention different during manual driving compared to driving with supervised automation?
3. Do these behaviours change with automation experience?

We studied automation use and driver visual attention allocation as a function of road type, speed, time in trip, time of day and time after first automation use. In order to perform these analyses, we explored to which extent the visual annotation of automation status and driver attention can be automated. We trained a classifier to identify system icons in the instrument panel using video and to classify driver attention distributions among attentive regions and regions associated with non-driving tasks using head pose estimated from video. Both classifiers were trained and evaluated on manually annotated data from the naturalistic study.

This study focuses on within-subject effects of automation use. We do not analyse differences or similarities between vehicle types, since they were not driven by the same participants or in the same conditions.

2. Methods

2.1. Data description

In a collaborative project conducted by TNO, SWOV and the Dutch ministry of Infrastructure and Watermanagement, the RDW (Dutch Vehicle Authority) and RWS (Dutch Road Authority), recent passenger cars with SAE level 2 automation were equipped with instrumentation to observe the driver and the environment. Naturalistic driving data was collected by providing these vehicles for daily use to drivers having no prior experience with SAE level 2 automation. The naturalistic dataset is unique in that it includes one month of manual driving (baseline condition) followed by two months of use with automation under naturalistic driving conditions (experimental condition), allowing for a longitudinal within-subjects analysis of how automation use changes over time. The full dataset includes five vehicle types (BMW 540i, Tesla S, Mercedes E, Volkswagen Golf E, Audi A4 Avant) driven by 20 participants. However, due to efforts needed for CAN data interpretation, the data from only two vehicle types (Tesla and BMW) and 9 participants could be analysed for this paper. An overview of the kilometres driven is provided in Table 1. For the Tesla drivers, data was collected successfully for 357 baseline and 431 experimental trips, while BMW drivers recorded 1044 baseline and 1387 experimental trips. For the remaining recordings automation status was either unavailable or inaccessible at the time of this analysis.

Both the BMW and Tesla were equipped with full-range ACC and LK. The BMW ACC operated for speeds between 0-180km/h while the Tesla ACC operated between 0-150km/h. In the BMW, LK permits hands off steering wheel for up to 25 seconds. While enabled, the BMW system provides supporting steering inputs whenever system requirements are met (e.g. clear lane markings) and allows the driver to provide corrective steering without disabling the automation. We refer to standby when it is enabled while operating conditions are not met. Tesla LK (at the time) permitted 15 seconds of hands free driving and becomes unavailable for the remainder of a drive when this limit is exceeded 3 times. Tesla's LK has to be engaged by the driver and turns off when the driver provides corrective steering or braking. The BMW allows LK use with or without ACC enabled. The Tesla only allows LK while ACC is on.

Table 1: Overview of the data collected.

	Baseline		Experimental	
	trips	days	trips	days
Tesla1	131	26	131	35
Tesla2	177	35	228	44
Tesla3	112	22	129	30
Total	402	83	488	109
BMW1	32	48	196	48
BMW2	111	34	154	50
BMW4	133	35	201	65
BMW5	62	21	109	36
BMW6	147	34	132	40
BMW7	1	1	137	46
BMW8	85	23	211	66
Total	571	196	1 140	351

Participants

For two participants (1 BMW, 1 Tesla), the demographic data was not available. The remaining 7 participants were all male, mean age 49 years (σ 5.2 years), licenced for 29.1 years (σ 6.2 years) and had driven 30,000km to 40,000km in the 12 months prior to the experiment. All participants indicated they felt “very interested" and “averagely" to “well" informed about the latest technological developments in the vehicle sector. Prior to the experiment, all but one participant normally used a vehicle equipped with cruise control, zero with adaptive cruise control or lane keeping assistance and three with lane departure warning. One participant (Tesla group) indicated to frequently use lane keeping assistance.

Instrumentation

Each vehicle was retrofitted with eight cameras observing the driver, instrument cluster, exterior in forward, left, right and rear directions, pedal bay and a top-down view towards the driver seat. The drivers were observed with 325x288 resolution at 10Hz. The Tesla instrument panel was observed with 720x576 resolution at 25Hz. Figure 1 provides an overview of the available video feeds. A smart camera system (MobilEye) recorded lane position and surrounding road users. For map-matching, GPS and IMU data were collected at 1Hz and 10Hz respectively.

CAN-bus data was collected, from which various signals were reverse-engineered, including velocity, accelerations, steering wheel angle and torque, brake and accelerator pedal, turn indicator, lights, wind screen wipers, and (for the BMW) status information on the automation and warning systems (lane departure; collision). For the purpose of this study, only velocity and automation status were used. All signals except video were time-stamped. Video recordings were not synchronised but were watermarked with a human-readable timestamp.

Figure 1: Overview of the eight camera perspectives recorded by the TNO instrumentation in the time-synchronized visualisation by SWOV for each vehicle. In reading order: right mirror view, forward view, left mirror view, driver face, instrument panel, rear view, driver seat, pedal bay. The driver's face is occluded for privacy reasons.

2.2. Data preparation

A number of challenges emerged after data collection. Reverse engineering of CAN bus data to identify automation status was successful for the BMW but not for other vehicle types. GPS tracking, used for obtaining road type data, was not always available with sufficient accuracy. Additionally, some videos were corrupted and had to be omitted from the analysis. Table 2 shows data availability after filtering, synchronisation and re-sampling.

Two data enrichment efforts were performed for the analysis in this study. The first was to retrieve Tesla automation status by automatic detection of icons in the video of the instrument panel. Details on the implementation, training and validation are available in Appendix A and obtained 99.33% accuracy.

The second enrichment aimed to automatically annotate driver attention from video. Head pose was inferred instead of driver gaze because we were unable to measure this reliably. Several studies have suggested that head pose can be an acceptable gaze substitute when classifying attention into relevant regions of interest. Lee et al. (2018) have demonstrated that attention classification from head pose is feasible for on-road driving and obtained classification accuracies in the order of 83% and higher. Similarly, Braunagel (2017) used head pose as a fall back for eyes-on-road classification when gaze data was unavailable. Henni et al. (2018) showed that eye based features and head based features can achieve a similar classification performance for on-road drowsiness detection. Further implementation and validation details are available in Appendix B. While we were able to reproduce the per-class performance reported by Lee et al. (2018), overall classification accuracy was 69% and intersections over union metrics were below 50%, which is insufficient for attention analysis. This suggests that inferring attention from head pose is not feasible for driving scenarios, and demonstrates the importance of using appropriate performance metrics to judge classifier performance with unbalanced data. Lacking the means to classify driver attention per region of interest, this paper uses head pose variance as indicator for *possible* changes in attention behaviour.

Table 2: Data fraction available after pre-processing.

	Tesla	BMW
Automation status	100%	61.8%
Speed km/h	80.3%	61.8%
Allowed speed km/h	65.9%	60.4%
Road type	63.4%	52.3%
Head pose	72.4%	56.6%

3. Results

3.1. Automation usage

For the Tesla drivers, there were 16 baseline trips with very brief moments (0.2% of time on highways) of ACC or ACC&LK use. These are attributed to status classification faults. For the BMW drivers, there were 15 baseline trips (10 by one participant, 4 by another) where some form of automation was used (55% of time on highways). Trips where automation was used during Baseline were excluded from analysis.

For automation use during the experimental condition, we first describe the distributions for both the Tesla and BMW drivers and then provide a statistical analysis. Automation status is observed with respect to road type, road speed limit, driving speed, time since the start of a trip and time of day.

During the experimental condition, Tesla users drove manually 55.8% of the time, 8.9% with ACC and 35.3% with ACC&LK. BMW users drove 52.8% manually, 0.9% with ACC, 2.6% with ACC + LK on standby (ACC&LK_{sb}), 27.9% with ACC + LK, 6.7% without ACC but with LK on standby (LK_{sb}), and 9.1% with LK active. Speed limiting was not used. Figure 2 shows automation use by speed limit and road type. For both vehicles, most driving time was spent on the highway, and ACC&LK was used most here (Tesla: 63.0% , BMW: 55.7%). Manual driving was however preferred when negotiating highway links. Automation was used very little on roads with speed limits below 70 km/h. In both vehicles, preference seems to be towards using ACC&LK over using either ACC or LK.

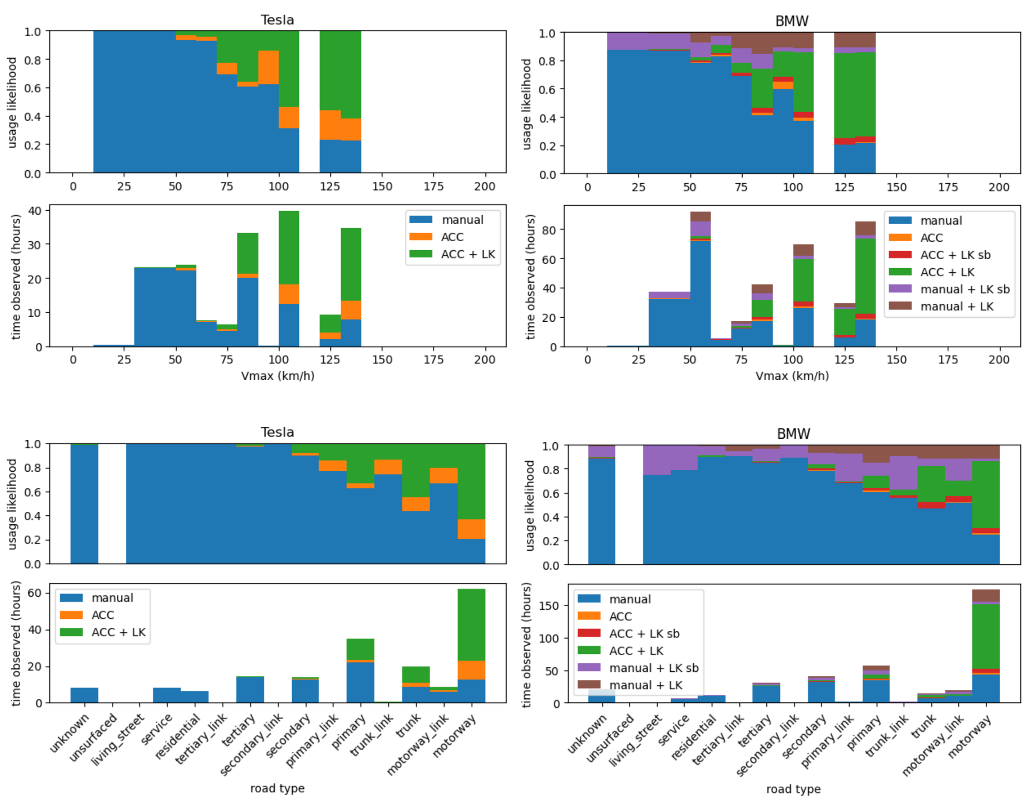

Figure 2: Automation use for road speed limit (top) and road type (bottom). Road type was obtained by map-matching using OpenStreetMap. LK_{sb} indicates lane keeping on standby.

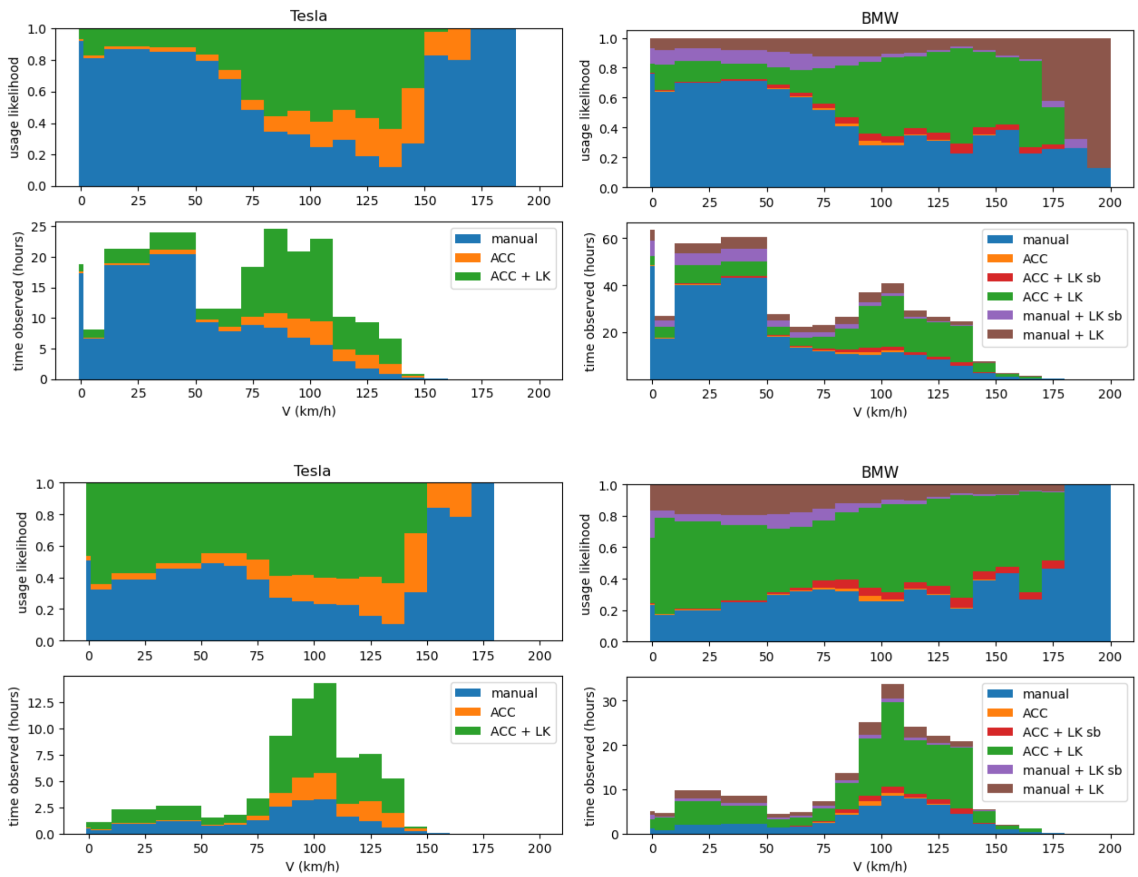

Figure 3: Automation use as a function of vehicle speed for all road types (top) and highways (bottom).

Figure 3 shows how automation use changes with driving speed. Usage was generally low for driving speeds below 70 km/h. However during highway driving, automation use remained high at all speeds, with peak usage during slow stop-and-go traffic (0-30 km/h) and higher speeds (>80 km/h). Drivers of the BMW quite often used LK with ACC off, especially at reduced speeds (30-80 km/h) on the highway. This suggests that longitudinal automation was not preferred or not trusted in dense traffic conditions, while LK was. This did not happen for the Tesla drivers, since LK is not available while ACC is off. At higher speeds a sudden drop in automation use can be observed. This drop corresponds with the upper limit at which the vehicles make automation available.

Figure 4 shows how automation use changes over the duration of a drive. After the first 10-20 minutes, automation use was relatively steady. The scatter at later times is an artefact resulting from the low number of long-duration trips. In the BMW data, a sudden drop in data availability occurs at 30 minutes. Since recordings are stored in 30 minute segments, some data loss may have occurred during these transitions. Figure 5 shows that automation use was uniform across the day for the Tesla drivers, while BMW users show more automated driving during commute hours (6h-8h and 16h-18h).

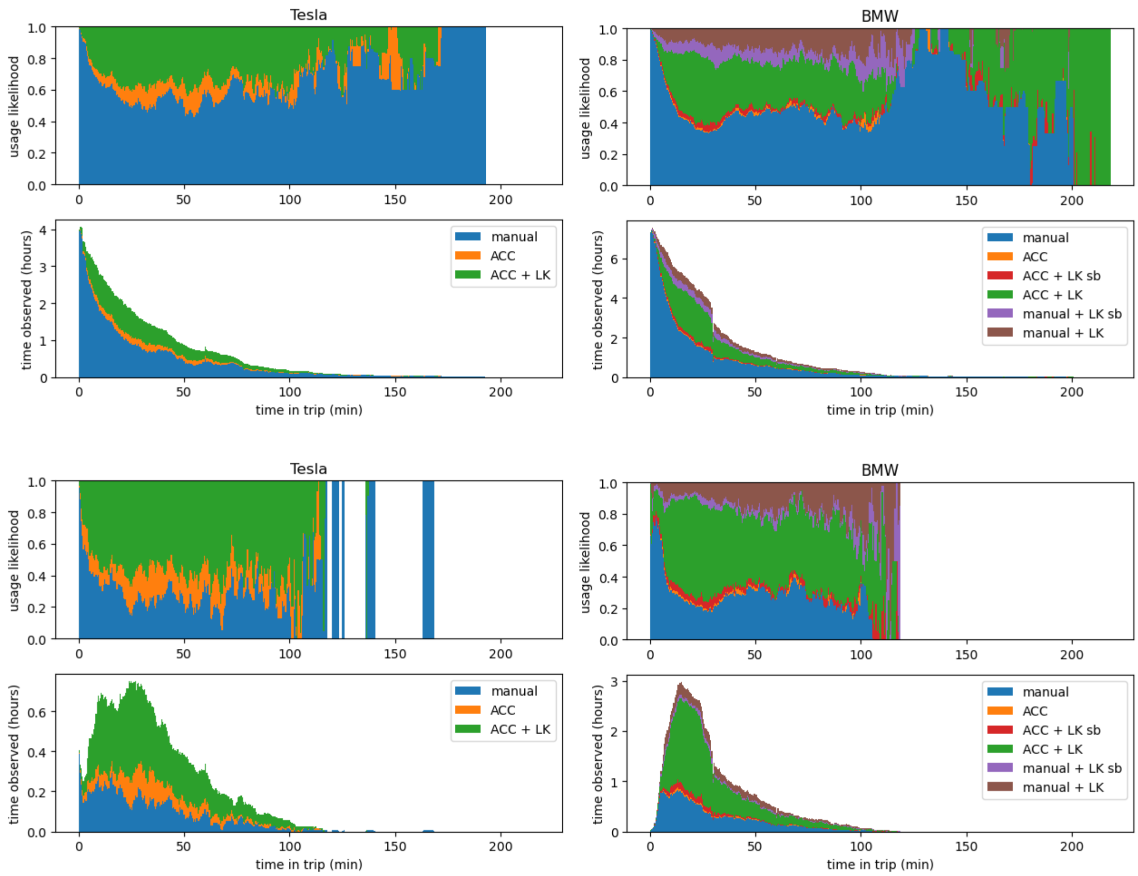

Figure 4: Automation usage over time since the start of a trip for all road types (top) and highways (bottom).

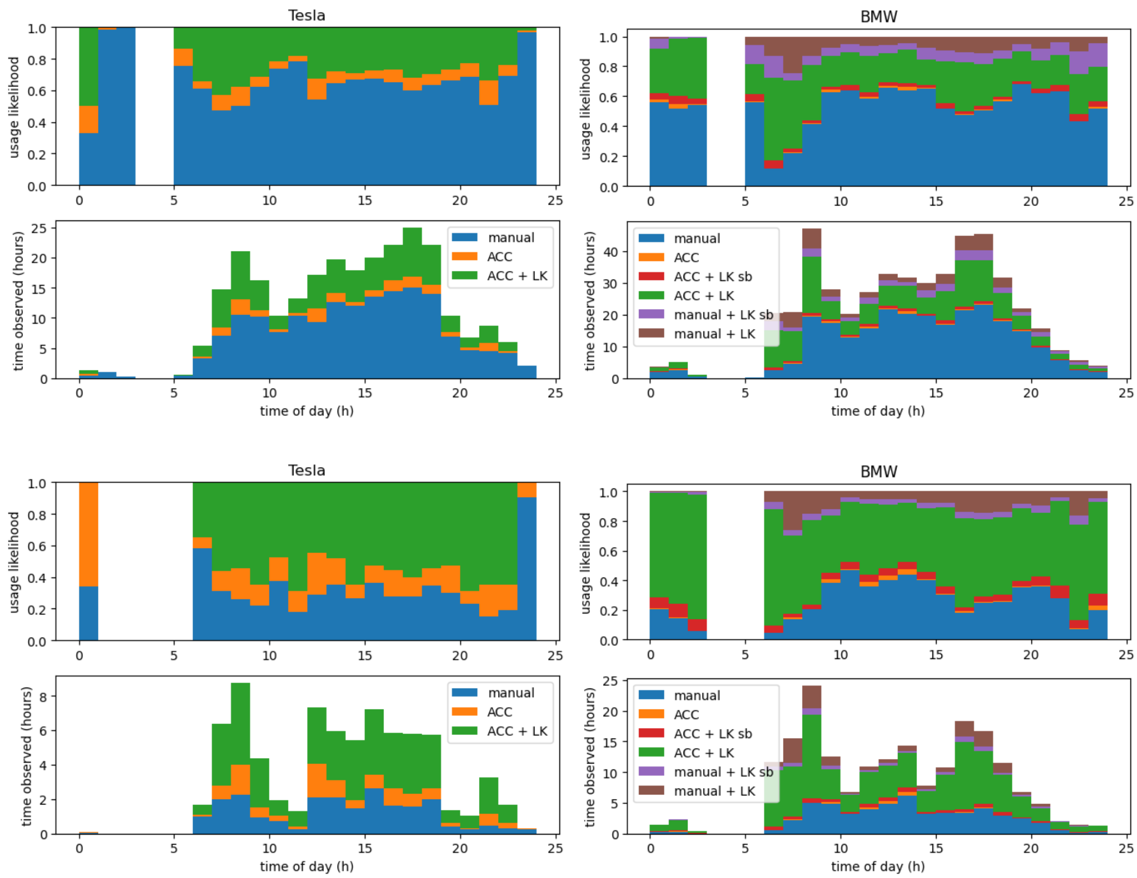

Figure 5: Automation use over time of day (Amsterdam DST) for all road types (top) and on highway (bottom).

Statistics of automation use

To evaluate if automation use was influenced by time in trip, time of day and driving speed, we performed between-trips multilevel ANOVAs with participant as random intercept variable. Only highway driving is considered for these analyses. Table 3 provides the means and standard deviations for each category and variable and the statistical results. For significant factors, effect sizes are presented as differences in estimated marginal means in Appendix D. It should be noted that Table 3 and the histograms of figures 5, 4 and in particular 3 show different distributions. This is because the histograms show total usage whereas Table 3 uses average usage per trip and does not account for trip duration.

Table 3: Descriptives and ANOVAs for automation use on highway for various effects over experimental trips. Note that night driving (23-4) is excluded from the ANOVA for the Tesla group because of small sample size.

			Time of day (hour)					Time in trip (min)			Speed (km/h)			
			23-4	5-9	10-15	16-18	19-22	0-30	30-60	60-90	0-10	10-60	60-100	>100
Tesla	nr. trips		3	45	97	60	34	212	103	43	95	155	160	155
	Manual	μ	72.9%	47.3%	57.1%	55.3%	47.4%	52.2%	36.5%	38.4%	74.1%	79.0%	42.1%	29.2%
		σ	34.7%	37.8%	37.6%	40.8%	39.7%	38.9%	35.2%	35.0%	40.1%	32.4%	34.1%	33.4%
	ACC	μ	27.1%	10.8%	9.9%	8.2%	10.5%	9.6%	17.7%	13.5%	3.5%	3.9%	13.9%	16.3%
		σ	34.7%	13.7%	14.7%	14.6%	17.5%	15.4%	26.4%	22.5%	16.8%	12.9%	19.9%	20.6%
	ACC&LK	μ	0.0%	41.9%	33.0%	36.5%	42.1%	38.2%	45.8%	48.2%	22.3%	17.1%	44.0%	54.5%
		σ	0.0%	32.9%	31.8%	35.6%	36.0%	34.4%	34.3%	35.4%	37.6%	29.6%	30.9%	33.6%
BMW	nr. trips		22	144	197	142	69	526	191	68	334	510	538	523
	Manual	μ	25.0%	22.4%	40.0%	29.4%	51.0%	34.3%	29.5%	34.6%	39.4%	48.2%	38.4%	31.0%
		σ	25.5%	33.6%	40.9%	39.1%	42.1%	39.8%	38.8%	42.2%	46.9%	46.0%	41.3%	39.9%
	ACC	μ	2.3%	1.2%	1.5%	1.1%	0.2%	1.1%	1.7%	1.0%	0.3%	0.4%	1.5%	0.9%
		σ	0.9%	6.6%	7.6%	5.8%	0.4%	6.5%	10.2%	4.6%	2.5%	3.1%	7.7%	6.1%
	ACC&LK^{sb}	μ	7.8%	5.1%	5.6%	4.2%	3.8%	5.0%	5.3%	3.2%	0.8%	2.1%	5.7%	5.8%
		σ	5.2%	6.0%	6.6%	5.8%	4.5%	6.5%	10.1%	5.4%	8.1%	10.3%	9.7%	8.3%
	ACC&LK	μ	60.5%	56.0%	42.3%	47.9%	34.8%	47.2%	44.7%	40.8%	21.3%	15.7%	38.8%	51.8%
		σ	26.2%	29.7%	32.8%	35.7%	33.7%	34.5%	32.0%	31.1%	36.4%	29.9%	33.1%	36.2%
	LK^{sb}	μ	1.9%	4.7%	4.1%	4.8%	3.9%	4.0%	6.6%	6.9%	22.4%	22.7%	6.7%	2.5%
		σ	3.8%	5.2%	8.6%	9.5%	8.5%	8.2%	16.5%	13.2%	39.3%	35.3%	12.5%	6.7%
	LK	μ	2.5%	10.7%	6.5%	12.5%	6.3%	8.3%	12.1%	14.0%	15.8%	11.0%	9.0%	8.0%
		σ	6.6%	15.8%	11.6%	21.0%	12.7%	15.8%	19.5%	17.6%	32.1%	21.5%	16.3%	17.0%

		Time of day		Time in trip		Speed	
		F	p	F	p	F	p
Tesla	Manual	F(3, 231.2)=0.603	0.613	F(2, 354.0)=5.257	0.006	F(3,559.2)=72.738	<0.001
	ACC	F(3, 231.4)=0.312	0.817	F(2, 355)=5.689	0.004	F(3,561)=18.937	<0.001
	ACC&LK	F(3, 231.3)=0.639	0.591	F(2, 354.0)=1.353	0.260	F(3, 559.2)=44.587	<0.001
BMW	Manual	F(4, 564.1)=5.530	<0.001	F(2,778.3)=9.015	<0.001	F(3, 1895.1)=24.039	<0.001
	ACC	F(4, 563.6)=0.883	0.473	F(2, 752.3)=0.406	0.666	F(3, 1897.2)=4.376	0.004
	ACC&LK$_{sb}$	F(4, 564.9)=2.7.55	0.027	F(2, 780.3)=3.514	0.030	F(3, 1895.5)=34.641	<0.001
	ACC&LK	F(4, 563.8)=3.574	0.007	F(2, 778.2)=0.268	0.765	F(3,1895.1)=150.453	<0.001
	LK$_{sb}$	F(4, 565.9)=0.205	0.936	F(2, 780.4)=11.459	<0.001	F(3, 1895.3)=29.808	<0.001
	LK	F(4, 564.2)=5.795	<0.001	F(2, 779.4)=4.451	0.012	F(3, 1895.2)=10.861	<0.001

Time in trip was split into three categories of 30 minutes each. For the Tesla group, the effect of time in trip was significant for ACC but not for ACC&LK use. Table D.9 shows that ACC use increased from 9.6% to 17.7% between the first and second 30 minutes of driving.

For the BMW drivers, time in trip did not change ACC or ACC&LK use significantly, but manual driving reduced by 12% in the second period of 30 minutes. This was replaced with 4.1% LK, 4.1% LK$_{sb}$ and 1.5% ACC&LK$_{sb}$.

Time of day was split into five categories: night (23:00 - 4:59), morning (5:00 - 9:59), day (10:00 - 15:59), afternoon (16:00 - 18:59) and evening (19:00 - 22:59). For the Tesla drivers, night time driving was omitted from statistical analysis due to low sample size. Effects of time of day on automation use were not significant. For the BMW drivers, automation use significantly changed with time of day for manual driving, ACC&LC_{sb}, ACC&LK use and LK use. There was no significant difference for ACC or LK_{sb}. Differences in estimated marginal means (Table D.8) suggest that ACC&LC_{sb} was used the most during night time driving, and was used 1.3% more during midday compared to evening commute. ACC&LK use did not differ significantly between times of day, with one exception being that it was used significantly less during evening drives (38.5%) compared to all other moments. Compared to midday drives, LK was used 5.1% more during morning and 7.0% less during afternoon commute hours.

Highway driving speed was divided into the same categories adopted by Naujoks et al. (2016). Driving speed had a significant effect on all forms of automation use in both the Tesla and BMW users. Estimated marginal means (Table D.10) show that manual highway driving occurred the most at speeds between 10-60 km/h for both vehicle types. Conversely, ACC&LK (and to a smaller extent ACC&LK_{sb} for the BMW) occurred the least at these speeds. ACC usage increased significantly over speeds between 10 km/h and 100 km/h for the Tesla drivers, but this trend was much smaller for BMW users. In the BMW group, LK without ACC was used significantly more while driving 10-60km/h compared to when driving 60-100 km/h, but not more compared to when driving >100km/h. Overall, the trend is towards more automation use (ACC or ACC&LK) at higher driving speeds. However, from a duration perspective, the overall ACC&LK usage in Figure 3 suggests that ACC&LK was used at lower speeds as much as at higher speeds. This may relate to different behaviour during short and long periods of slow highway driving. Prolonged low speed driving was rare; only 11% of trips with slow highway driving contained more than 3 minutes. This suggests that ACC&LK was especially used during longer periods of slow highway driving, and less when such speeds were only reached momentarily, for instance when entering or leaving a highway at slow speeds, or when traffic slowed down momentarily.

3.2. Attention distribution

Since driver attention classification was unsuccessful, we evaluated if automation use changed the head pose distribution. This can indicate when and to which extent automation use changes monitoring behaviour. Head heading and pitch distributions (Figures 6 and 7) were centred to the 50-percentile of each trip, and the standard deviation was compared across conditions. Statistical differences were explored during highway driving with a multilevel ANOVA using participant as a random intercept. For the BMW group, the standby variants ACC+LC_{sb} and LC_{sb} were excluded from this analysis. Table 4) shows that head heading and pitch were significantly affected by conditions in both vehicles.

Table 4: ANOVAs for effects of automation use on head heading and pitch deviation.

	Heading		Pitch	
Tesla	F(3,583.0)=12.243	<.001	F(3, 581.2)=8.412	<.001
BMW	F(4, 1595.4)=79.286	<.001	F(4, 1593.7)=70.542	<.001

Figure 6 shows that large heading angles generally occurred less during automated compared to manual driving, but this is mainly attributed to road type since heading distributions are more centered while on the highway. Figure 7 indicates that Tesla users tended to face up more and face down less while using automation, whereas BMW users tended to have a wider distribution of pitch angles while using automation compared to manual driving. It should be noted that these behaviours are not informative on where the driver is looking, as demonstrated in Appendix B. These effects reduce when only considering highway driving.

For highway driving, Table 4 indicates that both heading and pitch deviation differed significantly between automation use for both vehicle types. The large sample size allows for statistically significant results even if effect sizes (Table D.11) are small. The effects followed the same trends for the BMW compared to Tesla drivers. Head heading deviation was smallest during ACC use (Tesla 12.0°, BMW 4.7°) and largest while driving manually in the experimental condition (Tesla 15.7°, BMW 10.2°). Interestingly, heading deviation in the baseline period (Tesla 13.5°, BMW 9.5°) was significantly smaller, but did not differ significantly from ACC&LK. For BMW users, heading deviation was also significantly smaller during LK (7.9°) compared to baseline.

For both groups, head pitch deviation did not differ significantly between baseline (Tesla 6.6°, BMW 5.1°) and experimental manual driving and was significantly smaller during ACC (Tesla 5.6°, BMW 3.2°) compared to all other conditions. Pitch deviation during ACC&LK did not differ from the manual conditions (baseline and experimental) for the Tesla group, but was highest (5.6°) in the BMW group and significantly smaller (4.4°) during LK compared to the manual conditions, though the effect sizes are smaller than one degree (Table D.11).

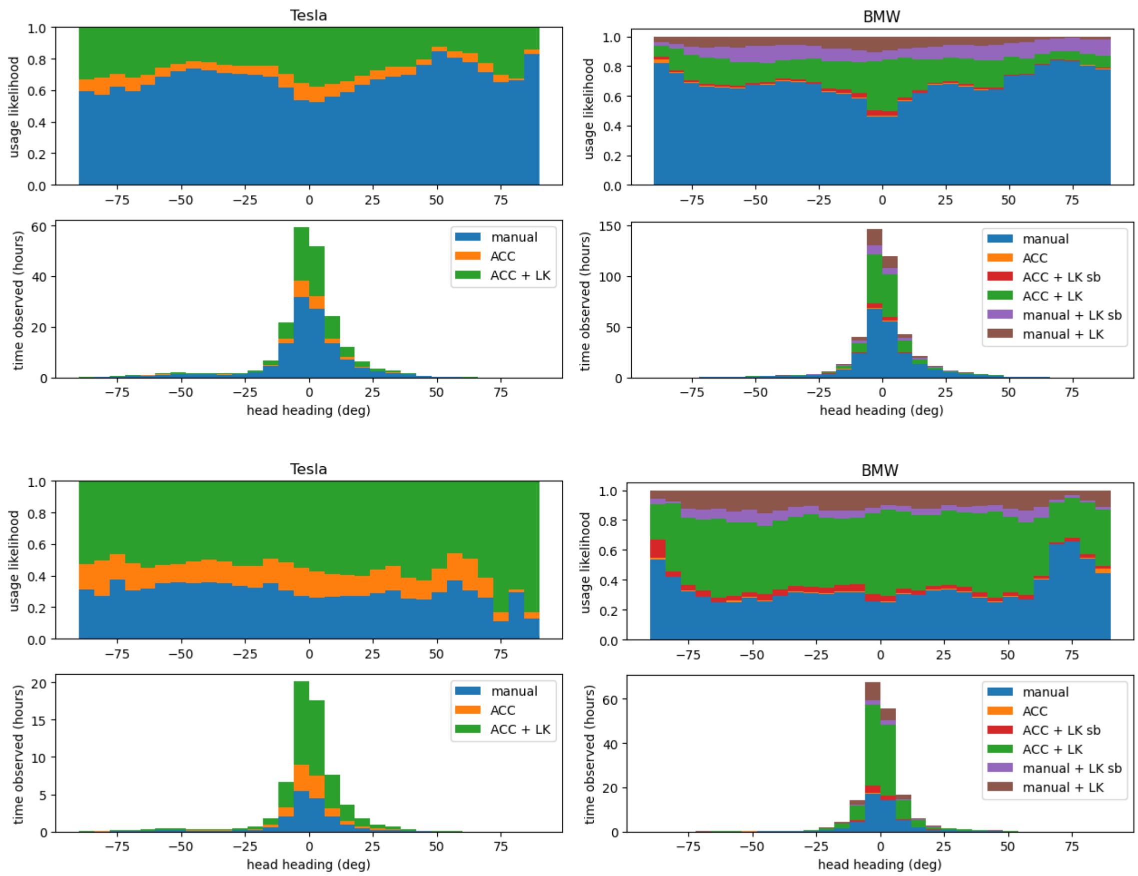

Figure 6: Distribution of head heading on all road types (top) and on highway (bottom). Positive heading indicates looking to the right and positive pitch indicates looking up.

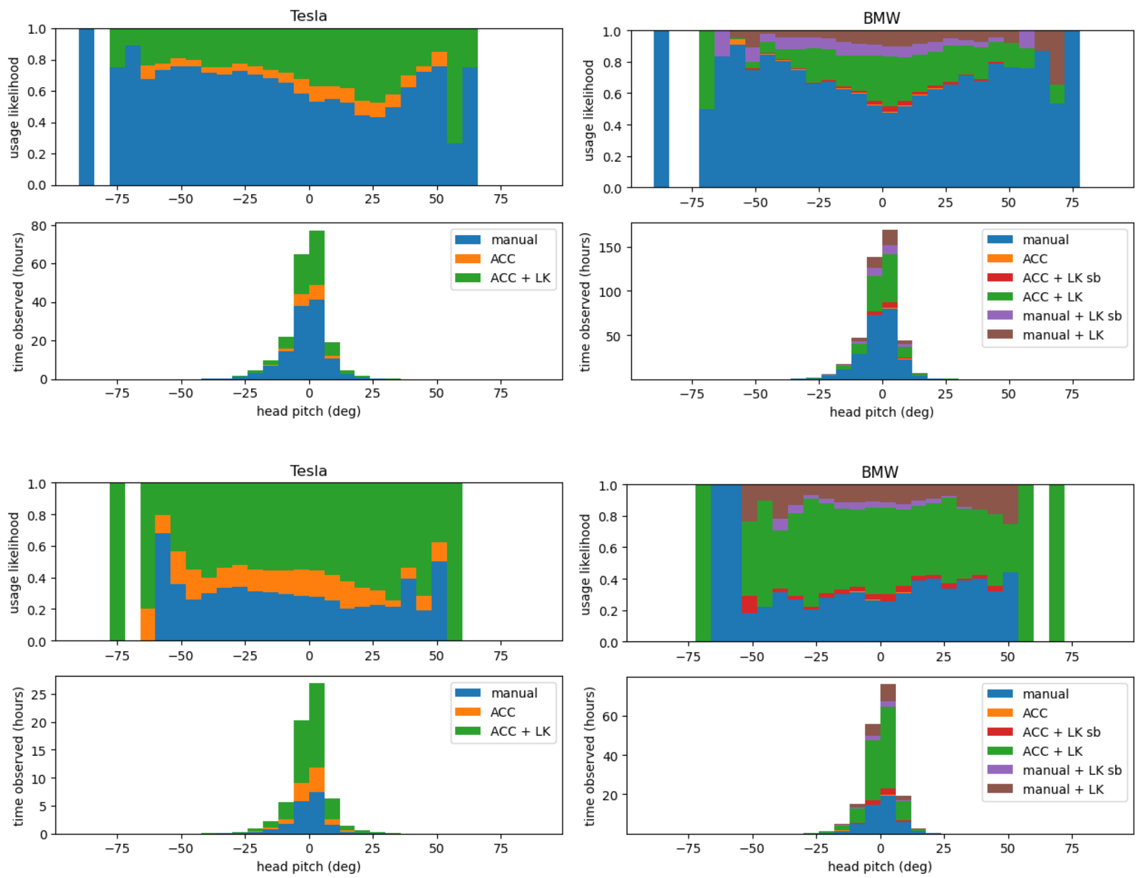

Figure 7: Distribution of head pitch on all road types (top) and on highway (bottom). Positive pitch is upward.

3.3. *Effects of experience*

We evaluate how automation experience during the first 2 months of the experimental condition changes automation usage, and if experience affects attention as indicated by head pose deviation. To accommodate the limited sample size, experience is examined in 3 week periods. The baseline period is included for manual driving. Statistics are in Table 5 and additional descriptives for automation usage over experience are given in Table D.13 which also includes the first day and first week of automation use.

There was no consistent change in automation usage over time. BMW users only had a significant change in ACC&LK and manual usage among the three-week clusters. Estimated marginal means suggest the effect is limited to weeks 6-9 where there was 9.5% less ACC&LK (p=.004) and 13.5% more manual driving (p<.001) compared to Wk

1-3 . Since this is the only effect observed, it is likely a consequence of uncontrolled differences between conditions, rather than a direct effect of experience.

Longitudinal changes in head heading and pitch deviation are examined as indicator for changes in attentive behaviour. Table 5 gives the main effects and Table D.12 provides pairwise comparisons for statistically significant effects.

For manual driving by Tesla users, head heading deviation was 3° higher throughout the experimental condition compared to baseline, but did not change significantly over time within the experimental condition. BMW drivers increased head heading deviation in manual driving only after 3 weeks of automation use, and only by 0.8°. During ACC, ACC&LK and LK use, head heading was not affected by experience for either vehicle type.

Head pitch deviation changed significantly over time for manual and ACC use among BMW users, and for ACC&LK use for Tesla users. For the Tesla group, head pitch deviation increased by 1.5° after the first 3 weeks of automation use while using ACC&LK. Pitch deviation during ACC among BMW users reduced 1.2° after the first 3 weeks of automation use, and during their manual drives increased 1.5° after 9 weeks of driving, which is likely an artifact of the reduced number of participants which participated that long. Similarly to the observed changes in automation usage, effects are inconsistent and occur at different times without explanation. Therefore, they are likely a consequence of uncontrolled differences between conditions, rather than a direct effect of experience.

Table 5: Main effects of experience (wk 1-3, wk 4-6, wk 7-9, wk 10-12) on usage and head heading and pitch deviation for highway driving in the experimental condition. ANOVAs are corrected for individual differences. Baseline condition is only included for Head variance during manual driving.

		Usage		Heading		Pitch	
		F	p	F	p	F	p
Tesla	Manual	F(3, 145.2)=0.241	0.868	F(4, 309.0) = 3.659	0.007	F(4, 316.1) = 1.329	0.259
	ACC	F(3, 156.0)=1.196	0.313	F(3, 129.0) = 2.032	0.113	F(3, 121.2) = 2.206	0.091
	ACC&LK	F(3, 114.3)=1.001	0.395	F(3,121.6) = 2.528	0.061	F(3, 128.0) = 7.611	<0.001
BMW	Manual	F(3, 513.7)=5.266	0.001	F(4, 628.6) = 15.234	<0.001	F(4, 627.9) = 2.687	0.030
	ACC	F(3, 516.9)=1.968	0.118	F(3,151.7)=2.112	0.101	F(3, 150.99)=4.812	0.003
	ACC&LK	F(3, 513.6)=2.851	0.037	F(3,379.6)=0.504	0.680	F(3, 379.4)=0.481	0.696
	LK	F(3, 515.4)=0.623	0.601	F(3, 342.4)=0.181	0.909	F(3, 341.5)=0.083	0.969

4. Discussion

This study analysed automation use across road types and speeds over the first two months of naturalistic use. Attention was evaluated using head pose deviation in heading and pitch.

4.1. Automation use

Across all road types, automation usage increased with speed for both vehicle types, with 92.5% manual driving at speeds below 70km/h for the Tesla users and 69.7% for the BMW users. The use of map matching for road type classification allowed us to analyse automation use per road type and discriminate low speed driving in congested highways from low speed driving related to the allowed maximum speed. Automation use differed substantially between these conditions, as illustrated in Figures 3 and 2. Automation (ACC and ACC&LK) was used most on highways, where it was used across all speeds including slow highway driving, but the least in congested traffic (10-60 km/h). Differences between total use time and per-trip averages suggest that manual driving is especially preferred when slow driving lasts shortly. This may include momentary slowdowns in traffic, but also transitions where the vehicle is entering or leaving the highway. Data of the BMW group suggests that ACC is used less in congested traffic, whereas LK is used. This suggests that unstable traffic flow impairs trust in longitudinal but not lateral automation performance. Otherwise, full automation (ACC&LK) is preferred over only using ACC for all road types, and speed limiting was never used. It is not clear if the latter means that ACC is preferred over speed limiting, since the participants have not experienced that system. This does not mean this system is not preferred in general; we can expect that OEMs know which features are desired among their user groups. These findings suggest that users were generally comfortable using automation during most highway conditions, and that automation was mostly used on road types for which the systems are intended. Automation use on urban roads was limited and incidental, which suggests that users are aware of the system's general limitations and typically act accordingly.

The effects of time of day did not show a clear circadian rhythm or other patterns. BMW drivers used ACC&LK least in the evening, most at night and did not differ significantly among morning, midday and afternoon drives. They used ACC more during morning and afternoon commute hours which could relate to road familiarity. Time of day effects were not observed for Tesla drivers. Time-in-trip effects suggest there was less manual driving after 30 minutes for both driver groups. However time-on-task effects did not occur for the dominantly used form of automation (ACC&LK) and effect sizes were small.

Similarly, no longitudinal experience effects on automation use were observed over the first 12 weeks of automation use. It is possible that most experience effects occur over a shorter period, possibly within a small number of trips. For instance, Beggiato et al. (2015) demonstrated that drivers converge a mental model of ACC within 3.5 hours of use. On the other hand, Larsson (2012) demonstrated that ACC users keep refining their awareness of system limitations over the first 10 months of use. It is possible that such adaptations are not expressed through overall metrics such as system usage.

4.2. Head pose deviation distributions

While we were unable to classify driver attention among attentive and driving unrelated areas, the analysis of head pose deviation identified small but significant trends in visual monitoring.

On the highway, head heading and pitch deviation were smaller during ACC use compared to other driving modes, including baseline. Deviations during ACC&LK did not differ from baseline manual driving, but heading deviation was larger during manual driving in the experimental phase compared to baseline driving. This difference between baseline and experimental manual driving could be caused by strategic automation use: drivers may prefer to drive manually in situations which require more head deviation, such as when changing lanes. Important to note is that the effect of ACC&LK on heading deviation depends on whether it is compared against baseline-manual (no difference) or experimental-manual (ACC&LK reduces heading deviation). This may raise caution for studies which compare attention between manual and automated driving without providing a manual baseline. Further research is needed to determine if this is an experience effect or a consequence of voluntary vs. instructed usage. Automation effects on head pitch deviation were very small and unlikely to carry practical significance.

Collectively, these findings suggest that the amount of attentional activity in terms of head pose deviation may be similar between ACC&LK use and baseline manual driving. This differs from Morando et al. (2019), who found that the median percent at road centre of glances was 3% smaller during SAE2 compared to manual driving. Possible explanations for this difference include the used metrics (gaze vs. head pose) and not controlling for periods of following a lead vehicle (which for Morando et al. increased percent road centre by 4%).

Besides the difference in head heading deviation between baseline and experimental manual driving, no longitudinal changes in head pose deviation were found which could indicate an effect of experience.

Whether the lower head pose deviation during ACC should be interpreted as an increase or decrease in monitoring intensity remains to be investigated. If drivers were mostly monitoring attentively during automation, lower deviation could indicate an increase in attention to road centre or cognitive narrowing due to an increased mental demand. However, it can also be caused by cognitive load from driving-unrelated thoughts (Victor et al., 2005; Wang et al., 2014), a reduced perceived need for visual scanning, or an increase in mind wandering (He et al., 2011). Even when gaze had been obtained in addition to head pose, identification of the correct cause may be challenging since even for gaze dispersion it is not certain if a wider deviation represents more distraction or a better monitoring strategy (Grüner and Ansorge, 2017). Classification of attention to driving related and unrelated areas may provide better insights.

5. Conclusions

5.1. When and in which conditions do drivers use ACC and LK support?

ACC and LK were mostly used on road types for which the systems are intended. On highways ACC&LK was used 63% of the time by the Tesla group and 57% of time by the BMW group. On highways, automation use generally increased with driving speed. It was used least in congested traffic (10-60 km/h) where ACC&LK was mainly replaced by LK (BMW) or manual driving (Tesla), which could mean that especially ACC is not preferred in unstable traffic. On urban roads and roads with speed limits below 70 km/h, automation was used less than 6% of the time, which suggests that users were aware of the system's general limitations in those conditions. Automation use was not clearly affected by time of day. Time-in-trip suggests that manual driving occurs less after 30 minutes of driving, but this did not lead to a significant increase in any of the separate forms of automation use.

5.2. Is driver attention different during manual driving and driving with supervised automation?

We found limited changes in monitoring behaviour with supervised automation. Head movement activity was smaller on highways compared to other road types. On highways, head pose activity during ACC&LK did not differ from baseline manual driving, but was smaller during ACC use. Head heading deviation was larger during manual driving in the experimental phase compared to manual driving in the baseline phase. This also means that studies can risk making incorrect inferences about automation effects on attention when only sampling manual and automation conditions during voluntary use without a baseline condition with instructed manual driving.

5.3. Do these behaviours change with automation experience?

There was no consistent change in automation usage over time. Similarly, changes in head motion activity could not be attributed to a simple experience effect, and are more likely a consequence of uncontrolled differences between conditions.

5.4. Limitations

This study includes 11 participants and two vehicle types. This sample size is too small to generalise the findings. Driver behaviour is generally more complex than what can be captured by the conditions examined here. Therefore, only the larger and consistent effects should be considered indicative. Another limitation is the use of head pose as indicator of attention. We demonstrated that driver head pose is not predictive of attended region of interest. While we argue that a change in head deviation can indicate a different monitoring strategy, we provide no suggestions on how such change should be interpreted with regards to better or worse monitoring, or its safety implications.

5.5. Suggestions for future research

Since few effects were observed for aggregate factors such as experience, time of day and time on task, future work could more closely examine motivations for automation use and disuse. Such information could be acquired through interviews with drivers or a close examination of the traffic situation when control transitions are taking place.

Our second recommendation is related to head pose. Since head pose tracking without gaze direction was insufficient for attention classification, we recommend gaze monitoring for future work on naturalistic attention monitoring.

Finally, for future research it would be interesting to study if different system interaction designs impact the effectiveness and usability of the systems, and how this differs across various user groups. Such insights could help formulate design choices that benefit safety and ease of use. Intuitiveness and ease of use of the systems are crucial for the adoption and safety of the system. Systematic evaluation can aid design guidelines for safe user interaction. Such guidelines would be of great value for both industry and policy. It would support industrial parties in designing safe and intuitive interfaces and it would support policy makers to evaluate new systems and set clear requirements for admission on their roads.

Acknowledgment This work was partially supported by the NWO-TTW Foundation, the Netherlands, under the project "From Individual Automated Vehicles to Cooperative Traffic Management - Predicting the benefits of automated driving through on-road human behaviour assessment and traffic flow models (IAVTRM)" -STW#13712.

The data collection was done by TNO with financial and in-kind contributions from the Dutch Ministry of Infrastructure and Water Management, Rijkswaterstaat, RDW, AON, PON Netherlands, BMW Netherlands and Germany, and Athlon Carlease. The research activities were conducted in collaboration with SWOV Institute for Road Safety Research at their premises in The Hague.

References

Ahlström, C., Victor, T., Wege, C., Steinmetz, E., 2012. Processing of eye/head-tracking data in large-scale naturalistic driving data sets. IEEE Transactions on Intelligent Transportation Systems 13, 553–564. doi:10.1109/TITS.2011.2174786.

Ahlström, C., Zemblys, R., Jansson, H., Forsberg, C., Karlsson, J., Anund, A., 2021. Effects of partially automated driving on the development of driver sleepiness. Accident; analysis and prevention 153, 16. doi:10.1016/j.aap.2021.106058.

Baltrusaitis, T., Zadeh, A., Lim, Y.C., Morency, L.P., 2018. Openface 2.0: Facial behavior analysis toolkit, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE. pp. 59–66. doi:10.1109/FG.2018.00019.

Banks, V.A., Eriksson, A., O'Donoghue, J., Stanton, N.A., 2018. Is partially automated driving a bad idea? observations from an on-road study. Applied ergonomics 68, 138–145. doi:10.1016/j.apergo.2017.11.010.

Beggiato, M., Pereira, M., Petzoldt, T., Krems, J., 2015. Learning and development of trust, acceptance and the mental model of acc. a longitudinal on-road study. Transportation Research Part F: Traffic Psychology and Behaviour 35, 75–84. doi:10.1016/j.trf.2015.10.005.

Braunagel, C., 2017. Ensuring the take-over readiness of the driver based on the gaze behavior in conditionally automated driving scenarios. Dissertation. Eberhard Karls Universität Tübingen. Tübingen. doi:10.15496/publikation-23076.

Dutch Safety Board, 2019. Wie stuurt? Verkeersveiligheid en automatisering in het wegverkeer. Technical Report. Onderzoeksraad voor veiligheid. URL: https://www.onderzoeksraad.nl/nl/page/4729/wie-stuurt-verkeersveiligheid-en-automatisering-in-het-wegverkeer.

Farah, H., Bhusari, S., van Gent, P., Mullakkal-Babu, F.A., Morsink, P., Happee, R., van Arem, B., 2021. An empirical analysis to assess the operational design domain of lane keeping system equipped vehicles combining objective and subjective risk measures. IEEE Transactions on Intelligent Transportation Systems 22, 2589–2598. doi:10.1109/TITS.2020.2969928.

Glaser, Y., Glaser, D., Green, C., Llaneras, R.E., Meyer, J., 2017. Driver hazard detection and avoidance performance as a function of eyes-off-road interval under partially automated driving. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 61, 1914–1918. doi:10.1177/1541931213601959.

Grüner, M., Ansorge, U., 2017. Mobile eye tracking during real-world night driving: A selective review of findings and recommendations for future research. Journal of Eye Movement Research 10, 1–18. doi:10.16910/jemr.10.2.1.

Hancock, P.A., Matthews, G., 2018. Workload and performance: Associations, insensitivities, and dissociations. Human factors , 374–392doi:10.1177/0018720818809590.

Harms, I.M., Bingen, L., Steffens, J., 2020. Addressing the awareness gap: A combined survey and vehicle registration analysis to assess car owners' usage of adas in fleets. Transportation Research Part A: Policy and Practice 134, 65–77. doi:10.1016/j.tra.2020.01.018.

He, J., Becic, E., Lee, Y.C., McCarley, J.S., 2011. Mind wandering behind the wheel: Performance and oculomotor correlates. Human factors 53, 13–21. doi:10.1177/0018720810391530.

Henni, K., Mezghani, N., Gouin-Vallerand, C., Ruer, P., Ouakrim, Y., Vallières, É., 2018. Feature selection for driving fatigue characterization and detection using visual- and signal-based sensors. Applied Informatics 5, 1. doi:10.1186/s40535-018-0054-9.

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

Large, D.R., Burnett, G., Salanitri, D., Lawson, A., Box, E., 2019. A longitudinal simulator study to explore drivers' behaviour in level 3 automated vehicles. 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '19 , 222–232doi:10.1145/3342197.3344519.

Larsson, A.F., Kircher, K., Andersson Hultgren, J., 2014. Learning from experience: Familiarity with acc and responding to a cut-in situation in automated driving. Transportation Research Part F: Traffic Psychology and Behaviour 27, 229–237. doi:10.1016/j.trf.2014.05.008.

Larsson, A.F.L., 2012. Driver usage and understanding of adaptive cruise control. Applied ergonomics 43, 501–506. doi:10.1016/j.apergo.2011.08.005.

Lee, J., Muñoz, M., Fridman, L., Victor, T., Reimer, B., Mehler, B., 2018. Investigating the correspondence between driver head position and glance location. PeerJ Computer Science 4, 19. doi:10.7717/peerj-cs.146.

Morando, A., Victor, T., Dozza, M., 2019. A reference model for driver attention in automation: Glance behavior changes during lateral and longitudinal assistance. IEEE Transactions on Intelligent Transportation Systems 20, 2999–3009. doi:10.1109/TITS.2018.2870909.

Naujoks, F., Purucker, C., Neukum, A., 2016. Secondary task engagement and vehicle automation – comparing the effects of different automation levels in an on-road experiment. Transportation Research Part F: Traffic Psychology and Behaviour 38, 67–82. doi:10.1016/j.trf.2016.01.011.

Park, H., Gao, S., Samuel, S., 2017. Modeling effects of forward glance durations on latent hazard detection. Transportation Research Record: Journal of the Transportation Research Board 2663, 90–98. doi:10.3141/2663-12.

Russel, S.M., Blanco, M., Atwood, J., Schaudt, W.A., Fitchett, V.L., Tidwell, S., 2018. Naturalistic study of Level 2 driving automation functions. Technical Report DOT HS 812 642. National Highway Traffic Safety Administration.

SAE, 2021. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Technical Report. SAE international. URL: https://www.sae.org/standards/content/j3016_202104.

Stapel, J., Mullakkal-Babu, F.A., Happee, R., 2019. Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. Transportation Research Part F: Traffic Psychology and Behaviour 60, 590–605. doi:10.1016/j.trf.2018.11.006.

Strickland, D.L., 2013. Visual-manual nhtsa driver distraction guidelines for in-vehicle electronic devices. National Highway Traffic Safety Administration 78. URL: NHTSA\T1\textendash2010\T1\textendash0053.

Sullivan, J.M., Flannagan, M.J., Pradhan, A.K., Bao, S., 2016. Literature review of behavioral adaptations to advanced driver assistance systems. AAA Foundation for Traffic Safety URL: https://aaafoundation.org/wp-content/uploads/2017/12/BehavioralAdaptationADAS.pdf.

Victor, T.W., Harbluk, J.L., Engström, J.A., 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. Transportation Research Part F: Traffic Psychology and Behaviour 8, 167–190. doi:10.1016/j.trf.2005.04.014.

Victor, T.W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., Ljung Aust, M., 2018. Automation expectation mismatch: Incorrect prediction despite eyes on threat and hands on wheel. Human factors , 1095–1116doi:10.1177/0018720818788164.

Wang, Y., Reimer, B., Dobres, J., Mehler, B., 2014. The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. Transportation Research Part F: Traffic Psychology and Behaviour 26, 227–237. doi:10.1016/j.trf.2014.08.003.

Appendices to be published on-line only.

Appendix A. Automation status for the Tesla group

Because CAN data was not successfully decoded for the Tesla, automation status was retrieved from the instrument cluster video through icon template matching and a simple neural network classifier. This approach was deemed infeasible for the other vehicles due to poor icon visibility in the recordings (which were challenging to discern even for manual annotation). The Tesla uses four icons to indicate system status: ACC-on, ACC-available, LK-on and LK-available. For each icon, three template images were selected to represent different light conditions and camera perspectives (which tended to change over the duration of the study). A confidence value for the presence of each template in the instrument cluster video was obtained every 12 frames (2.1 Hz) through OpenCV template matching performed on a 150 by 245 px search space to account for camera movement. These confidence values were presented to a simple neural network consisting of two hidden layers and leaky ReLu activation functions with a 0.1 negative slope. The full network and template icons (relative size as depicted) are illustrated in Figure A.8.

To train the classifier, 1628 status transitions were manually annotated among 121 randomly sampled recordings. This resulted in 206,653 frames for training. An additional test set with 445 transitions was annotated on 27 other videos, resulting in 108,396 frames for testing. Classification performance on the test set is shown in Table A.6 and resulted in an overall accuracy of 99.33%, which was considered sufficient for the current analysis. Since performance on the test set was used as a stopping criterion for classifier design efforts, performance of the final classifier was verified on another set of 69 randomly sampled videos through visual inspection. Among these, 1342 minutes of manual, 57 minutes of ACC and 257 minutes of ACC&LK use were observed. Approximately 3 minutes (0.2%) was mis-classified. Misclassification occurred when the icons were particularly challenging to detect from the video. Common artefacts include rolling image, occluding specular reflection and intensity overflow, as illustrated in Figure A.9. Specular occlusion typically resulted in momentary misclassification of a single frame. Pixel overflow could last for several seconds but was found to have negligible impact on classification performance. When rolling images occurred, they affected an entire recording. Based on frequency patterns in the classification, the 100 most suspect videos were manually checked for rolling image and 45 videos were removed from further analysis based on this check.

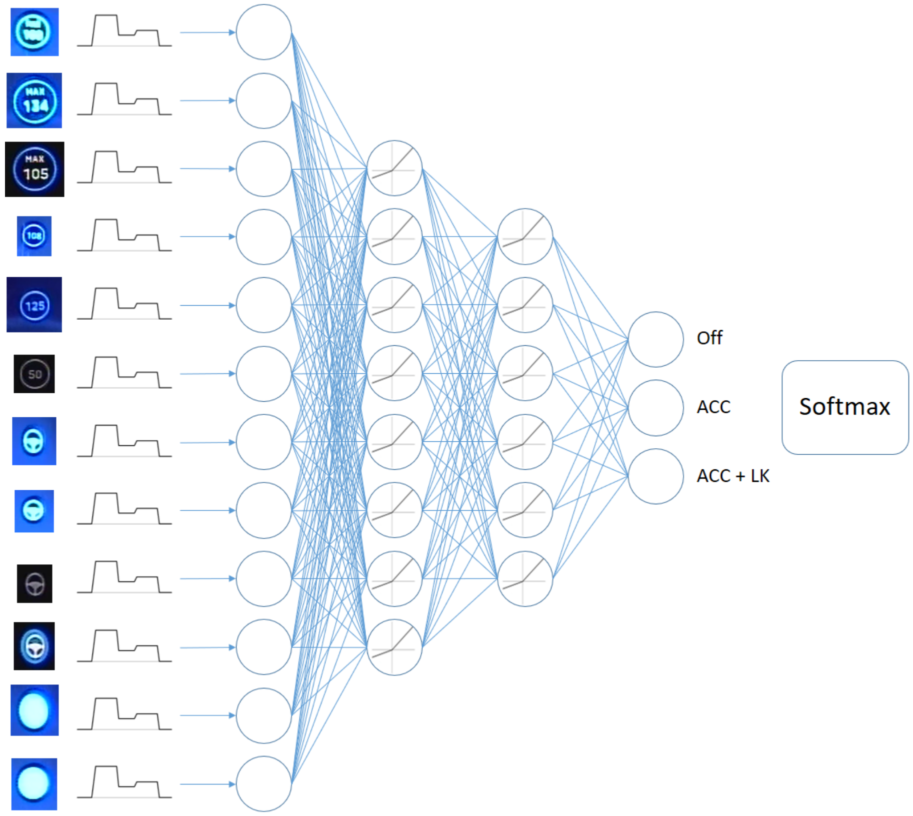

Figure A.8: Illustration of the classifier setup. Template matching was performed with 12 icon samples. The maximum normalized correlations were used as the input features of a neural network with two hidden layers (leaky ReLu activation functions with a negative slope of 0.1)

Table A.6: Confusion matrix comparing automation status between manual labelling and classifier on the test set. Accuracy is 99.33%.

		Predicted		
		Manual	ACC	ACC&LK
Annotated	Manual	54 347	55	100
	ACC	195	10 592	101
	ACC&LK	55	218	42 733

Figure A.9: Examples of glitches in video data on which icon recognition failed. Left: rolling image of the instrument panel as result of lost vertical sync and compression artifacts. Middle: specular reflection of a hand occludes the ACC icon. Right: byte overflow turns over-exposed white areas black in the in the icon.

Appendix B. Head motion tracking & regions of interest classification

To annotate driver attention behaviour, Head motion was tracked from the driver facing camera footage using OpenFace 2.0, an open-source facial behaviour analysis toolkit. It maintains a mean absolute error of 3° under various light conditions and facial expressions (Baltrusaitis et al., 2018). While OpenFace can also estimate eye gaze direction, it was found to perform poorly on the database and therefore not extracted.

Following promising results from Lee et al. (2018), We attempted to use head pose as substitute for gaze in our attempt to classify driver attention into relevant regions of interest.

We attempted to classify attention allocation in the Tesla to the regions of interest (ROI) defined in Figure B.10. Regions were selected for their functional purpose during driving; Left, right, windshield and instruments represent regions that are relevant for the driving task while distracted and centre console are not relevant to the driving or monitoring task.

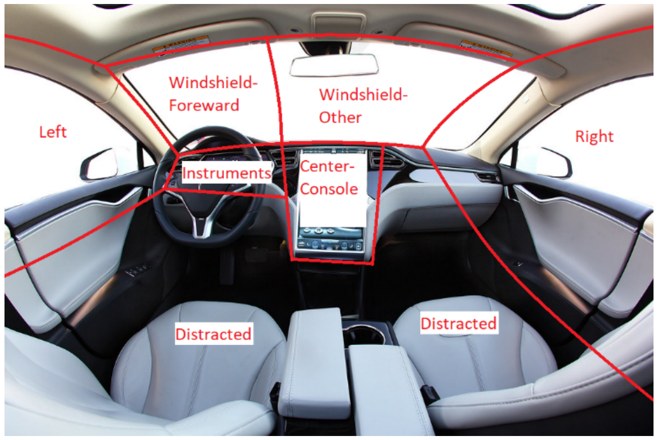

Figure B.10: Illustration of the approximate head pose regions of interest (ROI).

Since OpenFace estimates head pose relative to the driver-facing camera, association of these head poses to regions of interest requires calibration. Calibration had to be performed for each trip, because the camera tended to move over the duration of the study. Typical driver behaviour was used for calibration. A common approach is to create a histogram of all head poses and assume that the distribution modes (the most frequent direction) corresponds to facing the road centre on a straight road. One challenge with this approach is that it does not account for momentary postural changes, which may alter the relation between head pose and gaze direction. To account for this, Ahlström et al. (2012) identified multiple peaks as road-centre facing poses, amongst other refinements. In this study, we adopted a geometric solution in which the head heading and pitch are compensated for movement of the head's location. We determine the facing direction's intersection with a sphere with a 2 m radius centred at the 50-percentile head origin. This origin is determined for each trip and uses periods of highway driving if available, or all data otherwise. This intersection is then expressed in polar coordinates to retrieve a heading and pitch compensated for head origin. These angles can then be expressed relative to the forward facing reference angle, which we defined as the 50-percentile head pose while on the highway, or of the full data set when no highway data is available.

To create a ground-truth classification of head poses, we manually labelled 10,552 images from the driver-facing camera into six attentive and distractive regions, following the scheme in Appendix Appendix C. To balance the distribution of samples across regions, frames were sampled to obtain a uniform distribution of head poses. For the second half of the annotations, we filtered to only annotate stationary head poses since we found transitions between regions were often hard to classify. Since only very few poses were labelled as attending the instruments, this class was merged with windshield-forward, with which the samples overlapped best. Figure B.11 shows a scatter of all annotated head poses for sphere-projected heading and pitch angles.

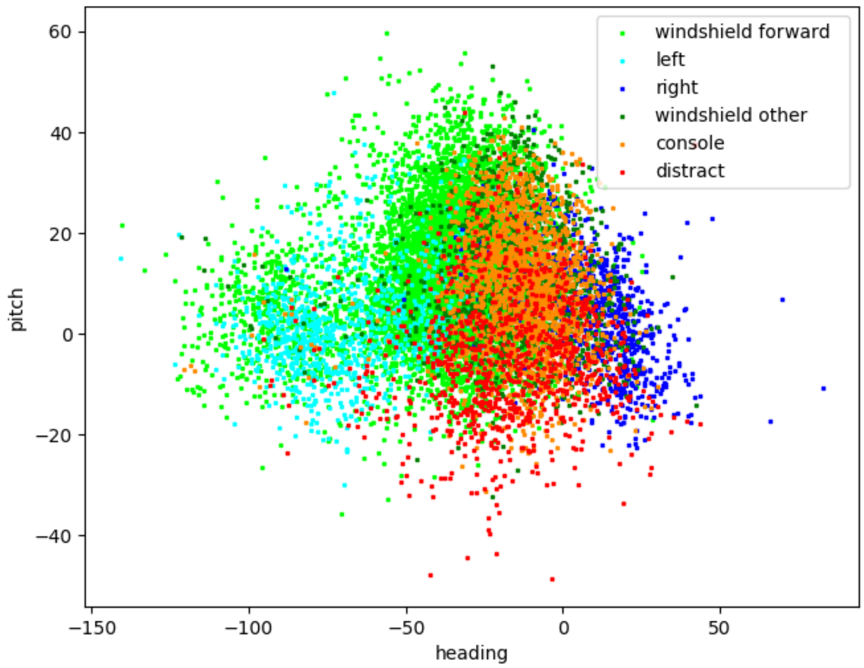

Figure B.11: Scatter of all annotated head poses. Values are compensated for head location through sphere projection, but orientations have not been corrected for camera placement. Hence zero heading is directed towards the camera and a heading around –35° is forward. Positive pitch represents facing upward.

Head pose was classified using a radial basis function support vector machine. 60% of the annotations were used for training and 40% for testing. Table B.7 shows the confusion matrix of the classifier on the test set. While accuracies between windshield-forward and other regions (Left: 92.7%, console: 86.4% and 95.7%) are similar to those reported by Lee et al. (2018) under similar conditions and methods, our overall accuracy is only 69%. Despite balancing head poses through uniform sampling, windshield forward received 6 times more annotations during manual labelling compared to the other categories on average. As a consequence, the classifier is biased towards this category and inflates accuracy for paired comparisons with that category. The intersection over union rates indicates performance without rewarding true negatives and thus provide a

Table B.7: Confusion matrix and intersection over union (IOU) for head pose classification, comparing human annotations to RBF SVM classifier over 4214 test images in the Tesla. Each cell indicates the number of images (top), percentage of ground truth annotated class (bottom left) and percentage of predicted class (bottom right)

		Predicted						
		Windshield forward	Left	Right	Windshield other	Centre console	Distracted other	IOU
Annotated	Windshield forward	2200	30	12	3	74	11	69.8%
	Left	152	124	0	0	1	2	39.9%
	Right	29	0	199	7	68	4	49.5%
	Windshield other	234	1	44	6	87	4	1.5%
	Centre console	318	1	23	11	285	29	29.8%
	Distracted	90	0	16	0	60	89	29.2%

better indication of classification performance per category. Even a binary classification between driving related and unrelated attention does not perform well. Grouping the interior-facing categories (distracted and centre console) results in an intersection over union of 41.3% which is insufficient for reliable distraction identification. The main source of confusion is the ambiguity between head facing direction and direction of gaze, which is especially large in pitch but also in heading for angles further away from road centre.

Due to the disappointing classification performance, the effects of naturalistic automation use on driver (in)attention distribution could not be analysed. Instead, head pose deviation is used as an indicator of attention.

Appendix C. Code book for annotating head orientation in the Tesla

Is frame suitable for annotation:

- Is driver present, and is driver's facing direction clear? No → Space
- Is another person's face clearly visible? Yes → Space
- Is driver's head pose hard to classify or exceptional/unconventional?∗ Yes → Space

If driver faces away:

- Is driver facing or glancing through left window or to left mirror? Yes → Left
- Is driver facing or glancing through right window? Yes → Right

- Is driver clearly looking away from any exterior view, mirror or vehicle display?∗∗ Yes → Distracted

If driver faces forward:

- Does driver glance to instrument panel? Yes → Instruments
- Does driver glance well below instrument panel? Yes → Distracted
- Does driver glance towards the rear view mirror? Yes → Windshield other
- Otherwise: Windshield forward

If driver faces towards the camera:

- Does driver glance (just) above the camera's origin? Yes → Windshield other
- Does driver glance at, slightly below or slightly right of camera?∗∗ Yes → Centre console
- Does driver glance slightly left and slightly below camera? Yes → Distracted
- Does driver glance through right half of windshield or right mirror? Yes → windshield other

∗ Examples include sneezing, being mid-motion and severe head tilt
∗∗ When holding nomadic device, consider the direction of attention rather than the activity.

Appendix D. Paired comparisons of estimated marginal means for significant effects

The following tables provide further details on the significance of differences between conditions for significant effects.

Table D.8: Differences in estimated marginal means for significant effects of **time of day** on highway **automation use**. Only shown for BMW as Tesla shows no significant effects. See Table 3 for descriptives.

		a - b	$\Delta\mu$ (a-b)	SE	p
BMW	Manual	5-9 h - 10-15 h	-10.5%	3.5%	0.003
		5-9 h - 16-18 h	5.1%	3.7%	0.167
		5-9 h - 19-22 h	-18.8%	4.6%	<0.001
		5-9 h - 23-4 h	4.4%	7.5%	0.558
		10-15 h - 16-18 h	5.4%	3.5%	0.125
		10-15 h - 19-22 h	-8.3%	4.4%	0.060
		10-15 h - 23-4 h	14.9%	7.4%	0.044
		16-18 h - 19-22 h	-13.7%	4.4%	0.004
		16-18 h - 23-4 h	9.5%	7.5%	0.207
		19-22 h - 23-4 h	23.2%	7.9%	0.003
	LK	5-9 h - 10-15 h	5.1%	1.6%	0.001
		5-9 h - 16-18 h	-1.9%	1.7%	0.270
		5-9 h - 19-22 h	3.9%	2.1%	0.065
		5-9 h - 23-4 h	4.2%	3.4%	0.219
		10-15 h - 16-18 h	-7.0%	1.6%	<0.001
		10-15 h - 19-22 h	-1.2%	2.0%	0.540
		10-15 h - 23-4 h	-0.9%	3.3%	0.782
		16-18 h - 19-22 h	5.7%	2.1%	0.007
		16-18 h - 23-4 h	6.0%	3.4%	0.077
		19-22 h - 23-4 h	0.3%	3.6%	0.933
	ACC&LK$_{sb}$	5-9 h - 10-15 h	-0.4%	0.6%	0.470
		5-9 h - 16-18 h	0.8%	0.6%	0.217
		5-9 h - 19-22 h	0.7%	0.8%	0.420
		5-9 h - 23-4 h	-2.9%	1.3%	0.026
		10-15 h - 16-18 h	1.2%	0.6%	0.043
		10-15 h - 19-22 h	1.1%	0.8%	0.155
		10-15 h - 23-4 h	-2.5%	1.3%	0.054
		16-18 h - 19-22 h	-0.1%	0.8%	0.857
		16-18 h - 23-4 h	-3.7%	1.3%	0.005
		19-22 h - 23-4 h	-3.6%	1.4%	0.026
	ACC&LK	5-9 h - 10-15 h	5.3%	3.0%	0.078
		5-9 h - 16-18 h	6.1%	3.1%	0.051
		5-9 h - 19-22 h	13.1%	3.9%	0.001
		5-9 h - 23-4 h	-4.4%	6.4%	0.490
		10-15 h - 16-18 h	0.9%	3.0%	0.769
		10-15 h - 19-22 h	7.8%	3.7%	0.036
		10-15 h - 23-4 h	-9.7%	6.3%	0.124
		16-18 h - 19-22 h	7.0%	3.9%	0.077
		16-18 h - 23-4 h	-10.5%	6.4%	0.100
		19-22 h - 23-4 h	-17.5%	6.7%	0.009

Table D.9: Differences in estimated marginal means for significant effects of **time-in-trip** on highway **automation use**. Time indicates duration into trip. See Table 3 for descriptives.

		a - b	$\Delta\mu$ (a-b)	SE	p
Tesla	ACC	0-30 min - 30-60 min	-8.1%	2.4%	0.001
		0-30 min - 60-90 min	-3.8%	3.3%	0.253
		30-60 min - 60-90 min	4.2%	3.6%	0.246
BMW	Manual	0-30 min - 30-60 min	12.2%	2.9%	<0.001
		0-30 min - 60-90 min	7.3%	4.4%	0.094
		30-60 min - 60-90 min	-4.8%	4.6%	0.293
	LK	0-30 min - 30-60 min	-4.1%	1.4%	0.004
		0-30 min - 60-90 min	-2.7%	2.1%	0.194
		30-60 min - 60-90 min	1.3%	2.2%	0.544
	LK^{sb}	0-30 min - 30-60 min	-4.1%	0.9%	<0.001
		0-30 min - 60-90 min	-4.4%	1.4%	0.002
		30-60 min - 60-90 min	-0.3%	1.5%	0.841
	ACC&LK^{sb}	0-30 min - 30-60 min	-1.5%	0.6%	0.016
		0-30 min - 60-90 min	0.4%	1.0%	0.640
		30-60 min - 60-90 min	2.0%	1.0%	0.050

Table D.10: Differences in estimated marginal means for significant effects of **driving speed** on highway **automation use**. See Table 3 for descriptives.

	a - b	BMW $\Delta\mu$ (a-b)	BMW SE	BMW p	Tesla $\Delta\mu$ (a-b)	Tesla SE	Tesla p
Manual	0-10 km/h - 10-60 km/h	-7.9%	2.4%	0.001	-4.4%	4.4%	0.324
	0-10 km/h - 60-100 km/h	2.9%	2.4 %	0.232	32.4%	4.4%	<0.001
	0-10k km/h - >100 km/h	10.0%	2.4 %	<0.001	45.3%	4.4%	<0.001
	10-60 km/h - 60-100 km/h	10.8%	2.1 %	<0.001	36.8%	3.8%	<0.001
	10-60 km/h - >100 km/h	18.0%	2.1 %	<0.001	49.7%	3.9%	<0.001
	60-100 km/h - >100 km/h	7.1 %	2.1 %	0.001	13.0%	3.8%	0.001
ACC	0-10 km/h - 10-60 km/h	-0.1 %	0.4 %	0.881	-0.3%	2.3%	0.886
	0-10 km/h - 60-100 km/h	-1.1 %	0.4 %	0.003	-10.3%	2.3%	<0.001
	0-10k km/h - >100 km/h	-0.6 %	0.4 %	0.133	-12.7%	2.3%	<0.001
	10-60 km/h - 60-100 km/h	-1.1 %	0.3 %	0.002	-10.0%	2.0%	<0.001
	10-60 km/h - >100 km/h	-0.5 %	0.3 %	0.127	-12.4%	2.0%	<0.001
	60-100 km/h - >100 km/h	0.5 %	0.3 %	0.105	-2.4%	2.0%	0.236
ACC&LK	0-10 km/h - 10-60 km/h	5.2 %	2.2 %	0.016	4.7%	4.2%	0.264
	0-10 km/h - 60-100 km/h	-18.7 %	2.1 %	<0.001	-22.1%	4.1%	<0.001
	0-10k km/h - >100 km/h	-31.5 %	2.1 %	<0.001	-32.7%	4.2%	<0.001
	10-60 km/h - 60-100 km/h	-23.9 %	1.9 %	<0.001	-26.8%	3.6%	<0.001
	10-60 km/h - >100 km/h	-36.7 %	1.9 %	<0.001	-37.3%	3.6%	<0.001
	60-100 km/h - >100 km/h	-12.8 %	1.9 %	<0.001	-10.6%	3.6%	0.004
LK	0-10 km/h - 10-60 km/h	4.7 %	1.4 %	0.001			
	0-10 km/h - 60-100 km/h	6.6 %	1.4 %	<0.001			
	0-10k km/h - >100 km/h	7.6 %	1.4 %	<0.001			
	10-60 km/h - 60-100 km/h	2.0 %	1.3 %	0.117			
	10-60 km/h - >100 km/h	3.0 %	1.3 %	0.018			
	60-100 km/h - >100 km/h	1.0 %	1.2 %	0.412			
ACC&LK^{sb}	0-10 km/h - 10-60 km/h	-1.1 %	0.6 %	0.081			
	0-10 km/h - 60-100 km/h	-4.8 %	0.6 %	<0.001			
	0-10k km/h - >100 km/h	-4.9 %	0.6 %	<0.001			
	10-60 km/h - 60-100 km/h	-3.7 %	0.6 %	<0.001			
	10-60 km/h - >100 km/h	-3.8 %	0.6 %	<0.001			
	60-100 km/h - >100 km/h	-0.1 %	0.6 %	0.803			
LK^{sb}	0-10 km/h - 10-60 km/h	-0.7 %	1.6 %	0.658			
	0-10 km/h - 60-100 km/h	15.1 %	1.6 %	<0.001			
	0-10k km/h - >100 km/h	19.3 %	1.6 %	<0.001			
	10-60 km/h - 60-100 km/h	15.8 %	1.4 %	<0.001			
	10-60 km/h - >100 km/h	20.1 %	1.4 %	<0.001			
	60-100 km/h - >100 km/h	4.2 %	1.4 %	0.003			

Table D.11: Differences in estimated marginal means for significant effects of **automation use** on highway **head heading and pitch deviation** during baseline and experimental conditions.

		$\Delta\mu$	SE	p
Tesla	Heading baseline - Manual	−2.2°	0.6°	0.001
	Heading baseline - ACC	1.5°	0.7°	0.025
	Heading baseline - ACC&LK	−0.3°	0.7°	0.682
	Heading Manual - ACC	3.7°	0.6°	<0.001
	Heading Manual - ACC&LK	1.9°	0.6°	0.003
	Heading ACC&LK - ACC	1.8°	0.7°	0.008
	Pitch baseline - Manual	−0.1°	0.3°	0.833
	Pitch baseline - ACC	1.1°	0.3°	<0.001
	Pitch baseline - ACC&LK	0.0°	0.3°	0.945
	Pitch Manual - ACC	1.1°	0.2°	<0.001
	Pitch Manual - ACC&LK	0.1°	0.2°	0.774
	Pitch ACC&LK - ACC	1.1°	0.3°	<0.001
BMW	Heading baseline - Manual	−0.7°	0.3°	0.010
	Heading baseline - ACC	4.8°	0.4°	<0.001
	Heading baseline - ACC&LK	<0.1°	0.3°	0.936
	Heading baseline - LK	1.6°	0.3°	<0.001
	Heading manual - ACC	5.5°	0.3°	<0.001
	Heading manual - ACC&LK	0.7°	0.3°	0.004
	Heading manual - LK	2.3°	0.3°	<0.001
	Heading ACC - ACC&LK	−4.7°	0.3°	<0.001
	Heading ACC - LK	−3.2°	0.3°	<0.001
	Heading ACC&LK - LK	1.6°	0.3°	<0.001
	Pitch baseline - Manual	−0.2°	0.1°	0.193
	Pitch baseline - ACC	2.0°	0.2°	<0.001
	Pitch baseline - ACC&LK	−0.4°	0.1°	0.001
	Pitch baseline - LK	0.7°	0.1°	<0.001
	Pitch manual - ACC	2.1°	0.2°	<0.001
	Pitch manual - ACC&LK	−0.2°	0.1°	0.046
	Pitch manual - LK	0.9°	0.1°	<0.001
	Pitch ACC - ACC&LK	−2.4°	0.2°	<0.001
	Pitch ACC - LK	−1.2°	0.2°	<0.001
	Pitch ACC&LK - LK	1.2°	0.1°	<0.001

Table D.12: Differences in estimated marginal means for significant effects of **experience** on **head heading and pitch deviation** on the highway. Baseline is only used for the manual conditions.

	Heading manual			Pitch ACC&LK		
Tesla	$\Delta\mu$	SE	p	$\Delta\mu$	SE	p
Wk1-3 - baseline	1.9°	0.7°	0.006			
Wk4-6 - baseline	3.0°	1.2°	0.011			
Wk7-9 - baseline	2.9°	1.1°	0.008			
Wk10-12 - baseline	0.8°	2.2°	0.716			
Wk4-6 - Wk1-3	1.1°	1.2°	0.367	1.5°	0.5°	0.004
Wk7-9 - Wk1-3	1.0°	1.1°	0.370	1.8°	0.4°	<0.001
Wk10-12 - Wk1-3	−1.1°	2.2°	0.605	1.2°	1.0°	0.232
WK7-9 - Wk4-6	0.1°	1.4°	0.961	0.3°	0.6°	0.580
Wk10-12 - Wk4-6	−2.2°	2.4°	0.353	−0.3°	1.1°	0.779
Wk10-12 - WK7-9	−2.1°	2.3°	0.364	−0.6°	1.0°	0.542

	Heading manual			Pitch manual			Pitch ACC		
BMW	$\Delta\mu$	SE	p	$\Delta\mu$	SE	p	$\Delta\mu$	SE	p
Wk1-3 - baseline	−0.1°	0.4°	0.714	−0.0°	0.2°	0.775			
Wk4-6 - baseline	0.8°	0.4°	0.020	0.1°	0.2°	0.693			
Wk7-9 - baseline	0.6°	0.4°	0.129	0.1°	0.2°	0.528			
Wk10-12 - baseline	7.6°	1.0°	<0.001	1.5°	0.5°	0.002			
Wk4-6 - Wk1-3	0.9°	0.4°	0.013	0.1°	0.2°	0.530	−1.2°	0.3°	<0.001
Wk7-9 - Wk1-3	0.8°	0.5°	0.095	0.2°	0.2°	0.425	−1.0°	0.5°	0.030
Wk10-12 - Wk1-3	7.8°	1.1°	<0.001	1.5°	0.5°	0.001	−0.8°	1.0°	0.412
WK7-9 - Wk4-6	−0.2°	0.5°	0.696	0.1°	0.2°	0.782	0.2°	0.5°	0.622
Wk10-12 - Wk4-6	6.8°	1.1°	<0.001	1.4°	0.5°	0.003	0.4°	1.0°	0.670
Wk10-12 - WK7-9	7.0°	1.1°	<0.001	1.4°	0.5°	0.005	0.2°	1.0°	0.851

Table D.13: Descriptives of automation use over experience during highway driving in the experimental condition. Automation use during manual baseline of the Tesla drivers is attributed to misclassification on the video processing. Wk 1-3 includes day 1 and Wk 1.

			Baseline	day 1	Wk1	wk 1-3	wk 4-6	wk 6-9	wk 9-12
Tesla		Nr. trips	401	17	105	256	133	69	30
	manual	μ	99.8%	76.5%	77.4%	79.9%	79.2%	72.8%	83.2%
		σ	1.5%	27.8%	26.5%	24.9%	25.4%	25.8%	26.8%
	ACC	μ	0.1%	7.1%	5.1%	4.3%	4.6%	3.5%	1.6%
		σ	1.0%	10.1%	8.7%	8.9%	8.4%	6.8%	3.9%
	ACC&LK	μ	0.1%	16.4%	17.5%	15.7%	16.2%	23.7%	15.2%
		σ	1.1%	23.2%	22.1%	20.6%	20.2%	22.5%	24.0%
BMW		Nr. trips	286	17	57	188	199	121	14
	manual	μ	98.1%	36.9%	19.3%	23.4%	28.7%	36.7%	53.8%
		σ	11.5%	45.7%	32.8%	36.2%	37.4%	43.9%	45.6%
	ACC	μ	0.5%	6.8%	0.9%	1.3%	2.3%	0.3%	0.2%
		σ	4.2%	16.9%	2.7%	5.7%	12.0%	1.5%	0.6%
	ACC&LK$_{sb}$	μ	0.2%	6.2%	5.5%	5.7%	5.2%	4.5%	3.6%
		σ	1.6%	7.5%	4.6%	5.3%	6.1%	7.0%	5.6%
	ACC&LK	μ	1.0%	40.1%	61.6%	58.2%	52.0%	50.7%	39.7%
		σ	7.9%	39.0%	31.7%	33.3%	34.7%	38.4%	39.5%
	LK$_{sb}$	μ	0.1%	2.5%	3.3%	2.8%	2.6%	1.7%	0.7%
		σ	0.6%	4.8%	7.1%	5.7%	8.2%	3.4%	2.6%
	LK	μ	0.2%	6.6%	9.4%	9.0%	9.2%	6.1%	1.9%
		σ	1.7%	13.8%	14.8%	15.2%	17.5%	13.0%	6.5%