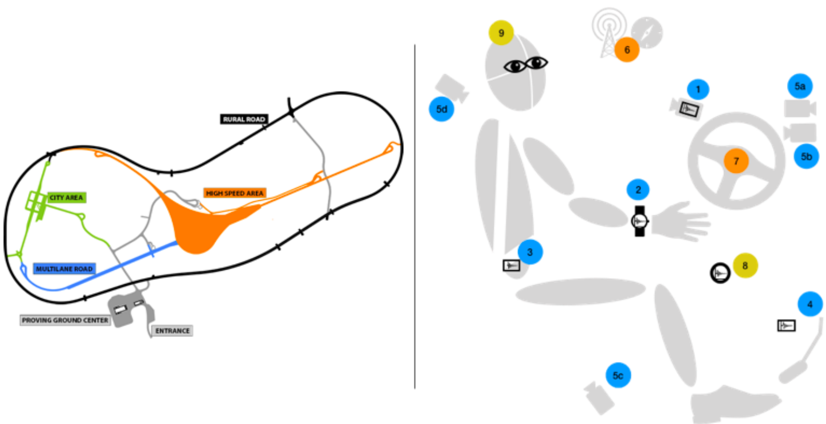
METHOD DEVELOPMENT FOR STUDIES AT TEST TRACKS OF LONG-TERM BEHAVIORAL ADAPTATION IN AUTOMATED VEHICLES

PROJECT REPORT
(A-0018)



Project funded by: SAFER Open Research at AstaZero (A-0018)
Authors: Andersson, J., Asker, E., Habibovic A., Klinegård, M., Lindström, D., Malmsten Lundgren, V.
Organization: RISE Viktoria AB
Project partners: Volvo Cars, Smart Eye
Contact: Azra Habibovic, azra.habibovic@ri.se
Date: 2018-06-30

Contents

Executive summary

To explore driver behavior in highly automated vehicles (AVs), independent researchers are mainly using driving simulators and conducting short experiments. This limits the ability to explore drivers' behavioral changes over time, which is crucial in the context of control transfer between drivers and AVs.

The project focused on a method development experiment that combines quantitative and qualitative data to capture effects of repeated interaction between drivers and AVs. Each driver (n=8) participated in the experiment on two different occasions (á 90 minutes) with one-week interval. On both occasions, the drivers traveled approximately 40 km on a rural road and encountered various traffic situations. They could engage automation or choose to drive manually. When activated, the automation was able to handle all situations on the road without requiring the driver to be engaged in maneuvering or monitoring (corresponding to SAE level 4). Examples of data collected include: vehicle state, driver biometrics, eye behavior, self-reported feeling of safety, as well as interviews and questionnaires capturing general impressions, trust and acceptance. The study was conducted at the AstaZero proving grounds in Sweden by Research Institutes of Sweden (RISE), Volvo Cars and Smart Eye.

The preliminary analysis shows that the first encounter effects were attenuated over time. The drivers went from being exhilarated (”fun”, ”cool”, ”impressive”) on the first occasion, to a more neutral state (”positive”, ”safe”, ”good”, “relaxed”) on the second occasion. Furthermore, there were smaller variations in drivers' self-assessed perceived safety at the second occasion, and drivers were faster to engage in non-driving related activities and become relaxed (e.g., they spent more time glancing off road and could focus more on non-driving related activities such as reading). These findings suggest that exposing drivers to AVs on two successive occasions may give more informative and realistic insights into driver behavior and experience as compared to only one occasion.

This study generated in-depth insights into behavioral changes of a limited number of drivers, however, these insights are difficult to generalize to a larger sample. Furthermore, the study was carried out during a winter period with varying adverse weather and road conditions, which may have affected driver behavior and embedded a bias in the findings. Our recommendation is that future studies of this type should be carried out under good weather and road conditions. Despite these challenges, the study generated knowledge on pros and cons with various data collection methods, something that is valuable for future studies.

The project was funded by SAFER Open Research at AstaZero, the Strategic Vehicle Research and Innovation (FFI) and the Knowledge Foundation (KK-stiftelsen).

1 Introduction

1.1 Background

Today, independent researchers use mainly driving simulators to explore driver behavior in highly automated vehicles (HAV, corresponding to SAE Level 4). These studies commonly last for 15-90 min. This limits the ability to explore drivers' behavioral changes over time, which is crucial in the context of control transfer between drivers and automated vehicles.

In this study, we explore methods for analyzing changes in drivers' behavior between two subsequent occasions. The focus is on automated driving on a rural road.

1.2 Aim

The overall aim of the study is to investigate user adoption effect of HAV, and what factors contribute to trust in HAV. This aim is broken down into the following two sub-aims.

- To develop a method that helps identifying after how long time a HAV driver begins to relax (trust automation) and how long it takes for the behavior to stabilize. Both qualitative and quantitative data will be collected and used to identify these behavioral changes.
- To explore if these behavioral patterns change between two subsequent test occasions (one week between occasions). In this way, we increase our knowledge of how much time a test participant needs to be exposed to a HAV to overcome the first encounter effect, which metrics can be used to identify the transition and whether it is possible to teach test participants to have a certain level of experience at recurring tests.

1.3 Research questions

- What factors indicate shift from first encounter (being tensed) to experienced (being relaxed)?
- How does driver behavior differ between the two occasions?

1.4 Research hypotheses

In this study, each driver participated at two different occasions. In each occasion, the drivers drove the test vehicle in manual driving (MD) mode or traveled in the automated driving (AD) mode, around the test track AstaZero several times (7 laps, corresponding to ca 40 km on rural road). They were exposed to similar conditions at both occasions. The major premise was that there would be a clear "first encounter effect" in driver behavior due to novelty of AD (i.e. drivers become tensed), and that this effect would reduce with the increased exposure to AD (i.e. drivers become relaxed as an indicator of trust). As such, we expected to see differences between laps, and between the two occasions.

The following hypotheses/assumptions were formulated to identify when the drivers become relaxed:

A. Drivers' **overall excitement** about the AD mode will decrease with time.
   - ➢ Driver is familiar with how the AD operates

B. When in the AD mode and relaxed, drivers will be faster to **engage in non-driving related activities** as compared to when in the AD mode and tensed.
   - ➢ Driver does not feel a need to monitor the AD

C. When in the AD mode and relaxed, **the ratio of glance time on road** will decrease as compared to when in the AD mode and tensed.
   - ➢ Driver does not feel a need to monitor the AD

D. When in the AD mode and relaxed, drivers' **self-assessed perceived safety** will be equal, or higher, than the corresponding value in the MD mode.
   - ➢ Driver trusts the AD

E. When in the AD mode and relaxed, **the foot hovering over the brake pedal** will decrease as compared to when in the AD mode and tensed.
   - ➢ Driver does not feel a need to be ready to take over the maneuvering control

F. When in the AD mode and relaxed, **the heart rate** will decrease as compared to when in the AD mode and tensed.
   - ➢ Driver is familiar with how the AD operates

G. When relaxed, **the handover of control to/from vehicle** will be experienced smoother as compared to when being tensed.
   - ➢ Driver is familiar with how the AD operates

2 Experimental design and execution

The experiment is a within-subject study. The independent variables include driving mode (MD/AD), and occasion (Occasion 1/Occasion 2), see Table 1[1].

Table 1 Conditions and levels

Independent variable	Conditions
Driving mode	Manual driving (MD) Automated driving (AD)
Occasion	Occasion 1 Occasion 2

The dependent variables include[2]:

- Time to engage in non-driving related activities
- Glance time off road
- Self-assessed perceived safety
- Foot on or hovering over the brake pedal
- Heart rate

[1] The original plan was to have an additional variable (environment type: test track/public road), however, it was omitted in the last minute due to adverse weather conditions.
[2] Several other variables were collected; however, these are not used here and are omitted from this report.

2.1 Test environment

The experiment was conducted at the test track AstaZero in Sandhult, Sweden. The test track consists of several traffic environments, one of which is the rural road where the experiment took place, Figure 1. It simulates a typical Swedish rural road setting. It consists of one lane in each direction separated only by a line. The posted speed limit is 70 km/h. The length of the road section is approximately 5.7 km (a loop).

Figure 1 An illustration of the test track AstaZero with the rural road with the different scenarios.

2.2 Data collection

The data were collected using several sources: a priori questionnaire, emotion assessing device, a posteriori questionnaire, biometric devices, eye-tracker and vehicle data log (Figure 2). These data are described in more detail in the following sections. It should be noted that only a portion of the data is analyzed and elaborated upon in this report.

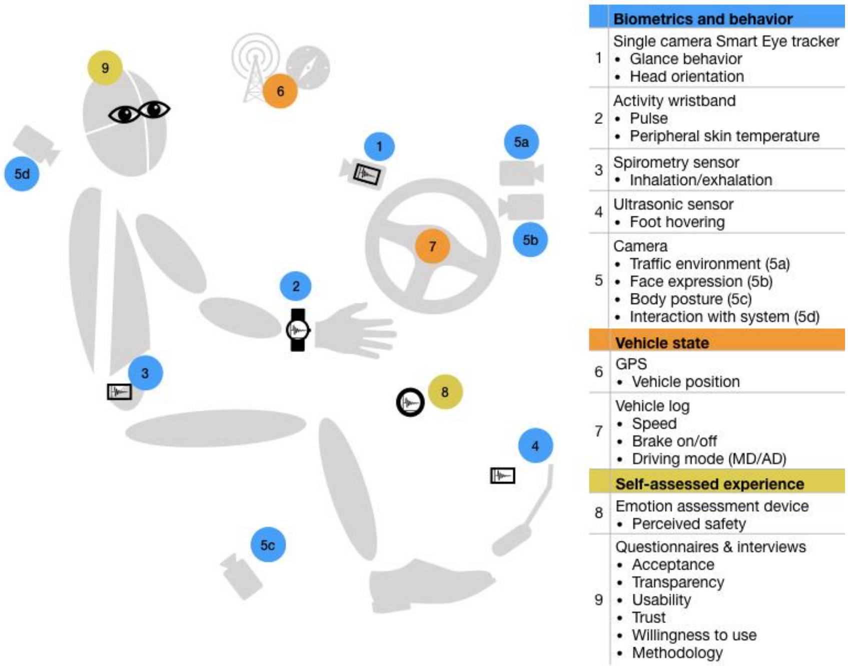


Figure 2 Sensor placement in the experiment vehicle and measurements.

2.2.1 A priori questionnaire

Each test participant (TP) was asked to complete a priori questionnaire (on paper). It captured the following information about the TP:

- Demographics (sex, age, etc.)
- Manual driving experience (number of miles driven)
- General opinion about ADAS
- Experience of using ADAS
- General opinion about HAV
- Experience of using HAV
- Familiarity with the test track AstaZero

2.2.2 A posteriori questionnaire

Each TP was asked to complete a posteriori questionnaire (on paper) that differed somewhat between Occasion 1 and Occasion 2. After the first occasion, the posteriori questionnaire consisted of questions derived from the extended Technology Acceptance Model (TAM)[3]. The questions were slightly modified to match the aim of this experiment. After the second occasion, the TP were asked to complete the same questionnaire as well as an additional TAM-based questionnaire where they compared their experiences and impressions from Occasion 1 and Occasion 2. In addition, they were asked to answer questions about the methodology, including the behavior of the automated function.

2.2.3 Emotions

After manual driving, the TP was asked to stop and assess his/her self-perceived safety using a specially designed emotion assessing device with a scale ranging from *"I'm not feeling safe at all"* to *"I'm feeling very safe"*. The scale contained a central value, see Section 2.6 for more details. When traveling in the AD mode, the TP was asked to assess his/her perceived safety using the same device (the vehicle did not stop).

2.2.4 Biometrics

The following biometric data were collected for each TP:

- Blood volume pulse (BVP), electro-dermal activity (EDA), also known as galvanic skin response (GSR), peripheral skin temperature and acceleration in 3 dimensions. The BVP measurement is used to calculate heart rate and heart rate variability (HRV).
- Inhalation/exhalation duration and slope, hold durations after inhale and exhale, inhalation/exhalation ratio, consistency and waveform morphology. From these measures, the person's mental state is derived, categorized as one of the cognitive states calm, tense and focus.
- Position of the foot relative to the brake pedal (cf. foot hovering).
- Head angle, gaze direction, pupil size, eyelid open/close ratio, and area of interest (i.e. the area where the TP is looking).

2.2.5 Vehicle log data

The experiment vehicle data that were logged included vehicle speed, GPS position, steering wheel angle, break force, AD state on/off, and turning indicators. In addition, the GPS position data of the vehicle involved in the traffic scenarios were also logged.

[3] Choi, J. K., & Ji, Y. G. (2015). Investigating the Importance of Trust on Adopting an Autonomous Vehicle. International Journal of Human-Computer Interaction, 31(10), 692-702. doi:10.1080/10447318.2015.1070549

2.3 Procedure

The experiment planning started in October 2017, and a first pilot was carried out at the test track AstaZero a few weeks later (Figure 3). The experiment was planned for November/December 2017, however, due to some practicalities, it was postponed to March 2018. Due to adverse weather conditions in March, an additional pilot was carried out immediately prior to the experiment. In total, six days at AstaZero rural road (and a garage) were used.

Figure 3 Experiment overall time plan.

On the experiment day, the test leader welcomed the TP at the AstaZero reception, and walked with the TP to a garage in the vicinity. In the garage, the test leader gave the TP introduction to the study and asked him/her to fill in a consent form and the *a priory* questionnaire (Figure 4). This was followed by an introduction to the test vehicle and equipment in front the garage. Next, the TP was asked to be seated in the vehicle and adapt seat and mirrors to fit his/her preferences. A calibration of the eye tracking equipment was then carried out with the help of an assistant, and the driving session could start.

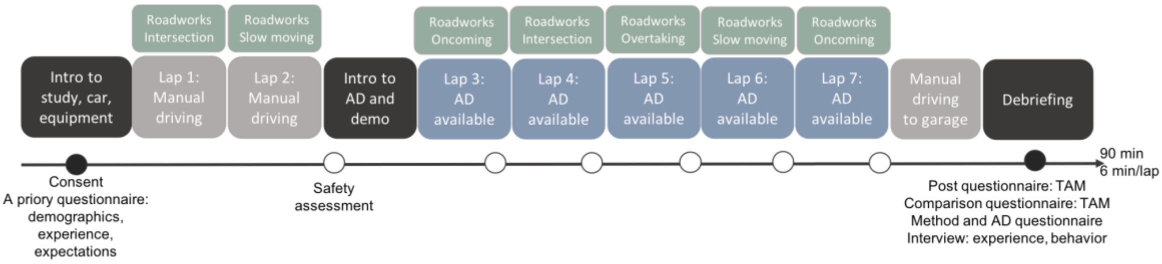


Figure 4 Experiment overview, including the scenarios that the test participant experienced in each lap.

In total, each TP travelled 7 laps on the test track rural road (ca 5.7 km). The test leader, assistant as well as a safety driver were seated in the backseat. The TP was first driving manually and after two laps, the TP was asked to stop and assesses his/her perceived safety by using an emotion assessment device in the vehicle. While in standstill, the assistant gave the TP an introduction to the automated driving (AD) function. The TP started the third lap by driving manually, and after a short while the AD became available. The TP could then choose to give the maneuvering control to the vehicle (i.e. activate the AD mode), and take back the maneuvering control at any time. While travelling in the AD mode (5 laps), the TP could engage in a self-chosen task (e.g., games, read newspaper, mobile surf). After each lap in the AD mode, the TP was asked to assess his/her level of perceived safety using the emotion assessment device (without stopping). The assessment was initiated at approximately same place using an audio signal. During the journey, the TP experienced the following traffic scenarios: roadworks, a slow-moving vehicle in the same lane / overtaking, a vehicle that is about to enter intersection, an oncoming vehicle (Table 2). Each TP experienced two scenarios per lap (Figure 4). The idea with these scenarios was to demonstrate for the TP the capabilities of the AD function. Each TP was exposed to the same experiment design and procedure at two different occasions (Occasion 1 and Occasion 2), with a week in-between. The experiment took about 90 minutes. The schedule is presented in Table 3.

Table 2 Several traffic scenarios were created at the test track to demonstrate the HAV capabilities.

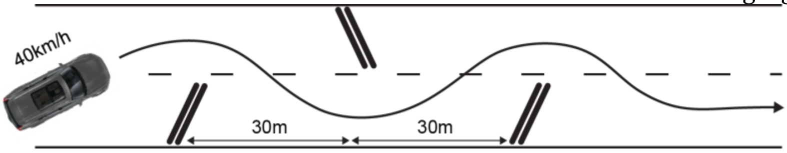
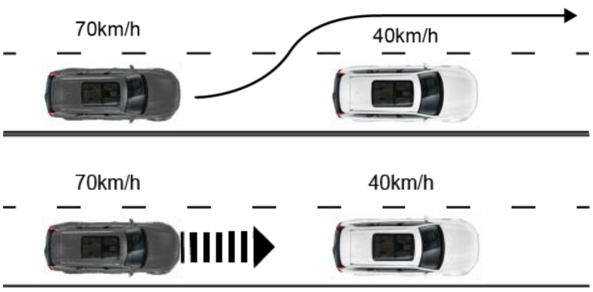
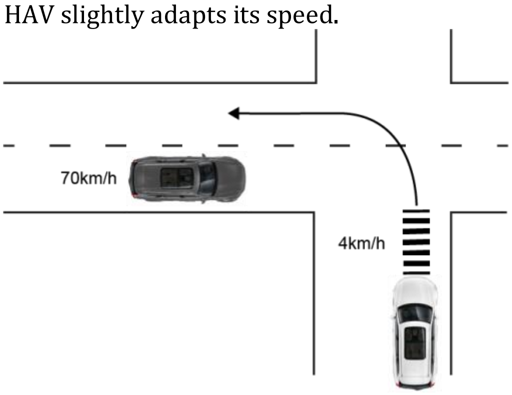
Scenario	Description
Roadworks	A static roadworks is marked by cones and a roadwork sign. It is visible from distance. The roadworks demonstrates that the HAV can handle zigzag driving. 
Upcoming & Overtaking	The HAV catches up a slow-moving vehicle (ca 40 km/h) in the same lane. The road section is straight, and visibility is good. The HAV decides to either follow the slow-moving vehicle (too short distance to intersection) or perform an overtake. This demonstrates that the HAV can handle overtaking. 
Intersection	A vehicle is about to enter the priority road (i.e. it should yield). Before entering it stops. To demonstrate that it has understood the situation and possible risk, the HAV slightly adapts its speed. 
Oncoming traffic	The HAV encounters a vehicle travelling in the adjacent lane in the opposite direction. 

Table 3 Schedule for the experiment on Occasion 1. A similar schedule was followed on Occasion 2.

Time (day 1)	TP #
10:30-12:00	TP 1
LUNCH	
12:30-14:00	TP 2
14:00-15:30	TP 3
15:30-17:00	TP 4

Time (day2)	TP #
08:30-10:00	TP 5
10:00-11:30	TP 6
LUNCH	
12:00-13:30	TP 7
13:30-15:00	TP 8
15:00-16:30	TP 9

2.4 Personnel

In total, six people were involved in the execution of the experiment: a test leader who was responsible for communication with and guidance of the TP outside the vehicle, an HMI assistant who was operating some parts of the in-vehicle HMI and guiding the TP inside the vehicle, a safety driver who was monitoring the automated driving and was ready to take over maneuvering control if needed, a driver responsible for operating the other vehicle that was used to create traffic scenarios, and two general assistants who were responsible for data collection and documentation.

2.5 Test participants

In total, eight (N=8)[4] test participants (TPs) were involved in the experiment (4 male and 4 female). They were recruited via an add on the RISE AB internal portal. To participate in the study, the TPs had to fulfill the following requirements: a) do not need glasses when driving[5], b) have a valid B driver license, c) can participate on two occasions, and c) is employed by RISE AB, and d) can participate during working hours. The condition d) was posed to ensure that the participants had appropriate insurances. In this case, RISE AB provided insurance to the employees even when taking part in the experiments. The TPs were engaged in the experiment voluntarily, and did not get any special personal benefits and payment. However, they could report their time (maximum 3h) on a project that was set up by RISE Viktoria.

2.6 Equipment

The following equipment was used in the experiment:

- Test vehicle (Volvo XC90)
- Scenario vehicle (Volvo XC90)
- 4 digital cameras (GoPro Hero 3+ and GoPro Hero 5)
- Eye-tracker (Smart Eye)
- Spirometry sensor (Spire Stone)
- Activity wristband (Empatica E4)
- Foot sensor (special made)
- Emotion assessing device (special made)
- Vehicle data logger (Vector FlexRay)
- iPhone for GPS logging of position in the scenario vehicle

These are described in more detail in the following sections.

[4] There was one more participant who participated in the pilot run.
[5] To enable eye tracking

2.6.1 Test vehicle

The experiment was conducted using a research Volvo XC90 vehicle that can be operated in manual driving (MD) mode and in automated driving (AD) mode. When the AD mode was activated, the vehicle was capable of handling the test track rural road with full autonomy and behaved as a level 4 vehicle according to the SAE-definition. Such a vehicle shall be able to handle all situations without any expectation on intervention from the driver. In this experiment, a safety driver positioned in the rear seat continuously monitored the automation and was ready to maneuver the vehicle.

2.6.2 Scenario vehicle

To create traffic scenario and demonstrate the capability of the HAV, a manually operated Volvo XC90 vehicle was used. It was equipped with the regular on-market systems.

2.6.3 Cameras

Four digital cameras of type GoPro Hero 3+ and GoPro Hero 5 were used to record behavior of the participants and the environment. The cameras were set to record video with 1920x1080 pixel resolution at 30 frames per second. The field of view (FOV) setting was set to 'Wide'.

2.6.4 Eye-tracker

A single camera Smart Eye system was mounted on the top of the instrument panel in front of the driver and used for eye and head tracking. It recorded different measures related to a TP's head and eyes such as head angle, gaze direction, pupil size and eyelid open/close ratio.

2.6.5 Spirometry sensor

A Spire Stone sensor measures the movements of the TP's torso to calculate measures related to respiration - inhalation/exhalation duration and slope, hold durations after inhale and exhale, inhalation/exhalation ratio, consistency and waveform morphology. From these measures, the TP's mental state is derived and categorized in three cognitive states: calm, tense and focus. The physical activity of the TP is also categorized in two states: active and sedentary. However, the activity state is not used in this study. The average breathing rate and the corresponding cognitive state were reported once per minute.

2.6.6 Activity wristband

An Empatica E4 activity wristband measures different physiological state variables of a person using sensors mounted on a wristband. The sensors are used to measure blood volume pulse (BVP), electro-dermal activity (EDA), also known as galvanic skin response (GSR), skin temperature and acceleration in 3 dimensions. The BVP measurement is used to calculate heart rate and heart rate variability (HRV). The Empatica E4 wristband device also contains a memory for the collected data and two channels for communicating collected data to external systems - a USB port and Bluetooth wireless communication radio. The skin temperature and acceleration measurements made by the Empatica E4 wristband are not used in this study.

2.6.7 Foot sensor

An ultra-sound sensor is used to measure the position of TP's foot relative to the brake pedal (foot hovering). It is integrated into a purpose-built multi-sensor device that is based on an Arduino[6] and used to record and log measurements to an memory card (SD). The device has a GPS unit used for high-accuracy time information.

2.6.8 Emotion assessing device

The purpose-built multi-sensor device is also connected to a panel (emotion assessing device) used for collecting a subjective rating of the participant (corresponding to a Likert scale 1-100). This panel consists of a small loudspeaker, a control knob and a push-button. The position of the knob is continuously recorded and logged. The scale of the knob ranges from "*I'm not feeling safe at all*" (left) to "*I'm feeling very safe*" (right) with a central value denoted by a single line. The loudspeaker emits a gentle beeping sound when the pushbutton is pressed. It was used to ask the TP to assess his/her safety.

2.6.9 Vehicle data logger

Data from the test vehicle (e.g., speed, indicators) were recorded in real-time using a Vector FlexRay data extracting device. These data were sent to an on-board computer and then via a wireless connection to a PC, where they are logged.

2.7 Data analysis

Given the fact that the study involved only 8 test participants, it was not feasible to conduct any in-depth statistical analyses. As such, we mainly report the frequencies and trends in the data. However, when feasible we use non-parametric statistics.

The focus is on identifying how TPs change their behavior with the increased number of laps in the AD mode, and how this behavior compares to their behavior in manual driving mode. Another focus is on identifying differences between Occasion 1 and Occasion 2.

3 Results

In the experiment, several data sources and data types were explored. In this report, a portion of the collected data is analyzed and elaborated upon. The rest of the data are to be analyzed yet. It should also be noted that we have considered TP1 is a pilot run and excluded it from further data analysis.

3.1 First impression of the automated function

This section relates to the assumption A: Drivers' overall excitement about the AD mode will decrease with time.

The interview analysis shows that the first encounter effects were attenuated over time. The drivers went from being exhilarated (”fun”, ”cool”, ”impressive”) on the first occasion, to a more neutral state (”positive”, ”safe”, ”good”, "relaxed") on the second occasion (Table 4).

The analysis of how the TPs reflected upon their behavior indicate that they first possessed a first encounter behavior in terms of curiosity (looking around, testing, learning) during Occasion 1. In Occasion 2, the TPs reported feeling more safe, relaxed,

[6] An open-source electronics platform based on easy-to-use hardware and software.

having more trust towards the system. To exemplify, one of the TPs described his/her experience in the following way:

> *"...one need to practice to lose attention, it is easy to start monitoring traffic even if you trust the car"* (Occasion 1, TP3)
> *"...last time I knew very little about the car and how to handle it...it was more exciting last time, but now it was a more pleasant experience"* (Occasion 2, TP3)

Table 4 The TPs answer to the first interview question: What is your spontaneous reaction to the automated function?

Test participant (TP)	Expression Occasion 1	Expression Occasion 2
TP2	Fun, exciting	Fun, nice
TP3	Wonderful	Safe
TP4	Passed my expectations	Positive, safe
TP5	Surprised	Positive
TP6	Fun, impressive	Good, it worked
TP7	Cool, interesting, impressive	Good
TP8	Positive, passed my expectations	Good, relaxed
TP9	Wonderful, it worked well	Good, it behaves expectedly

3.2 Engagement in non-driving activities

This section relates to the assumption B: When in the AD mode and relaxed, drivers will be faster to engage in non-driving related activities as compared to when in the AD mode and tensed.

The TPs were asked to bring gadgets that they would like to use while traveling in the AD mode. A few newspapers and magazines were also available in the vehicle. Here, we interpret the TPs' engagement in a non-driving related activity as a sign of relaxation and feeling safe enough to disengage from continually monitoring the vehicle itself while traveling in the AD mode.

A Wilcoxon signed-rank test showed that there is a statistically significant difference (p = 0.029, alfa=0.05) between Occasion 1 and 2 in terms of reduced time to when the TPs engage in non-driving related activities. This difference is illustrated in Figure 5, while Table 5 shows individual differences for each TP and occasion. As it can be seen, for some TPs there are relatively small differences between Occasion 1 and 2, yet the results still point towards a notable behavior change between the occasions.

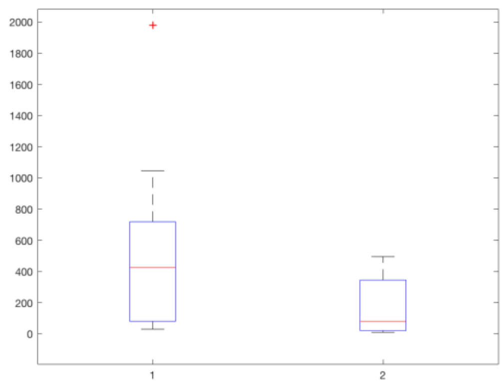


Figure 5 Time (in seconds) to engaging in a non-driving related activity in Occasion 1 (left) and Occasion 2 (right)

Table 5 Time to engaging in a non-driving related activity.

TP	Occasion 1 (seconds)	Occasion 2 (seconds)	Difference between the occasions (seconds)
2	1045	20	-1025
3	30	60	+30
4	335	80	-255
5	1980	95	-1885
6	425	480	+55
7	95	20	-75
8	35	10	-25
9	490	300	-190

The analysis of self-chosen non-driving related activity indicates that most of the TPs engaged in reading newspaper/magazine (TP2, TP7, TP8, TP9) on Occasion 1, while on Occasion 2 most of them interacted with their mobile phone (TP2, TP3, TP4, TP6, TP8, TP9). Even though the TPs performed similar activities on both occasions, the interviews showed that there was a difference in the level of engagement (i.e., the TPs' ability to disengage themselves from the driving task). On Occasion 2, the TPs stated that they could engage more deeply in their chosen activity as well as faster:

> *"…it felt like I could trust it faster since I've done this once before, when I started reading [on smart-phone] I almost forgot that I was in a self-driving car, I was totally into the reading..."* (Occasion 2, TP3)

> *"...I picked up the newspaper quickly, I felt that I know now how it works..."* (Occasion 2, TP7)

These differences between the two occasions are also evident from the post questionnaire that the TPs completed after Occasion 2. In Figure 6, the questions have been grouped to highlight the behavioral change, and lack of change, over time. The questions on ability to relax and being more willing to engage in other activities in the vehicle stand out as having changed over time; all TPs reported that they could relax more quickly and were more willing to engage in other activities during Occasion 2 as compared to Occasion 1. This reveals a first encounter effect in Occasion 1 that has to some extent been reduced with increased experience of the automated function. Also, some participants reported higher mental effort during Occasion 1, which fits well with the increased relaxation during Occasion 2 (Figure 6).

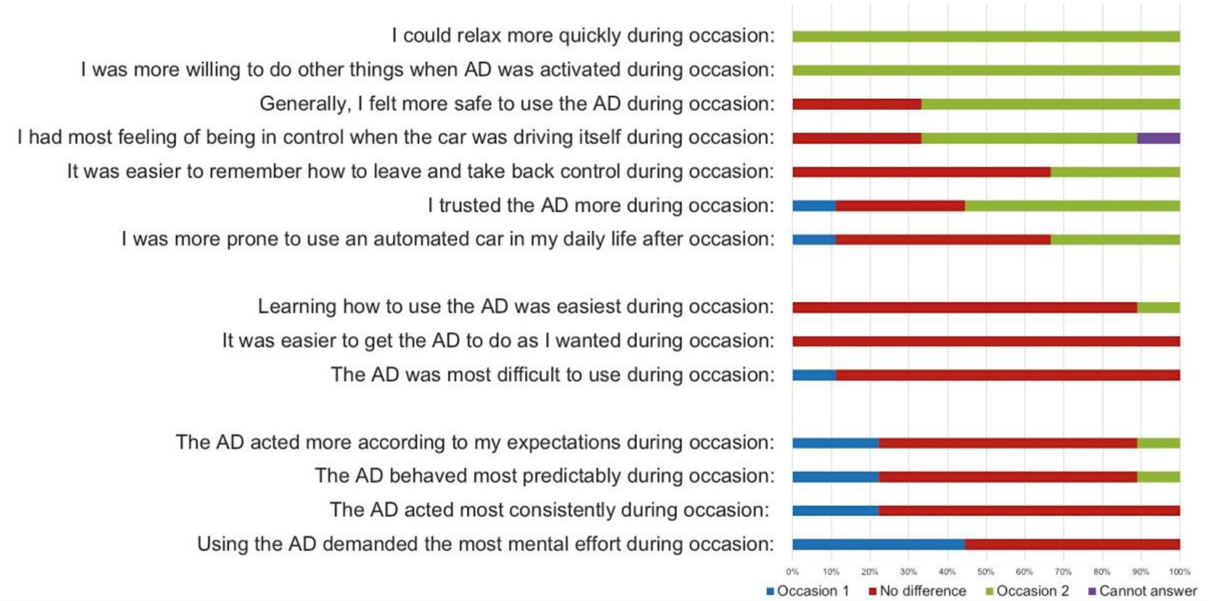


Figure 6 The test participants' self-assessment of behavioral change between Occasion 1 and Occasion 2.

3.3 Eyes off road

This section relates to the assumption C: When in the AD mode and relaxed, the ratio of glance time on road will decrease as compared to when in the AD mode and tensed.

The eye tracking data was processed to identify where the TPs were looking. Figure 7 shows the different target areas that were pre-defined to enable mapping of where the TP is looking. Here, we elaborate on data from Occasion 1. The analysis of data for Occasion 2 and comparison with Occasion 2 is to be done yet.

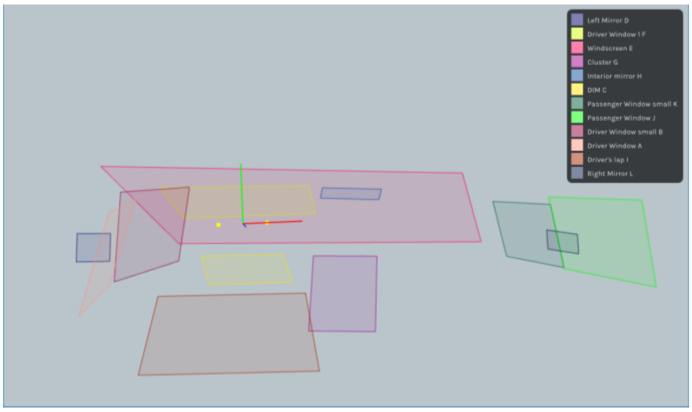


Figure 7 Eye tracking 3D world model of car interior.

The analysis of data from Occasion 1 shows that for most TPs there is a trend that the time spent watching through the windows decreases over time. At the same time, there is a trend that the time that the eye tracker failed to track the gaze, or TP looked in a direction where no target had been defined, increased over time. An example is shown in Figure 8 that illustrates the relative amount of time that the TP 4 looked a) through a window, b) down and c) the time that the eye tracker failed to track the gaze or if TP 4 looked in a direction where no target had been defined. A detail investigation of video recordings for TP 4 shows that the increase of c) was due to the increasing engagement in phone and magazine reading. Altogether, this indicates that the trust to automation for TP 4 increased over time. However, a high ratio of gazing through the windows does not necessarily mean that the person does not trust automation. We noticed that several TPs were watching the surroundings, or just looked up to get a feeling of orientation.

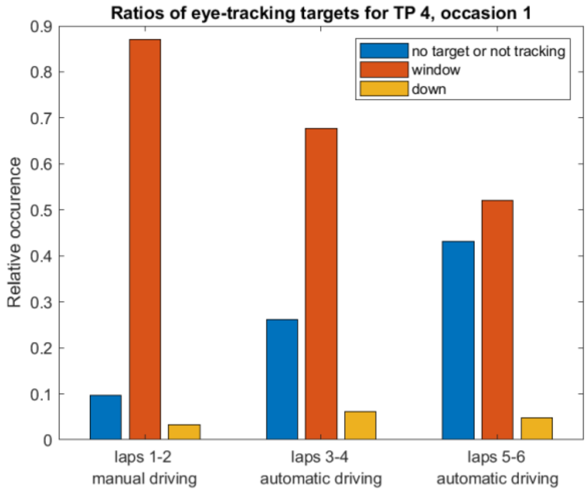


Figure 8 Relative amount of time spent on different target areas for TP 4.

3.4 Self-assessed perceived safety

This section relates to the assumption D: When in the AD mode and relaxed, drivers' self-assessed perceived safety will be equal, or higher, than the corresponding value in the MD mode.

Perceived safety was assessed using emotion assessment device in the test vehicle. Self-assessment was made after lap 3 (i.e. after finishing manual drive), and after each lap in the AD mode.

In Table 6, we can see that for Occasion 1, the perceived safety drops after the first lap in the AD mode as compared to the MD mode for all TPs except one (TP 8). Over time, the perceived safety seems to increase for most of the TP. For Occasion 2, there is a smaller dip after the MD mode, the trend of increased feeling of safety is not as clear as on Occasion 1, and the variation is smaller (Figure 9). The reduced variation in ratings could indicate that participants are getting used to the automated function. However, external factors such as adverse weather conditions during Occasion 2 (slippery road and snow) probably influenced ratings, at least for the manual driving.

Table 6 Self-assessed perceived safety for each TP along with mini-diagrams illustrating the trend.

	Occasion 1							Occasion 2						
	Man	Lap3	Lap4	Lap5	Lap6	Lap7		Man	Lap3	Lap4	Lap5	Lap6	Lap7	
TP2	100	48	59	63	69	72		89	77	75	86	84	87	
TP3	89	77	88	73	81	90		66	67	72	75	79	79	
TP4	89	79	76	82	85	85		84	77	81	81	81	81	
TP5	92	91	94	95	96	98		86	88	90	90	90	92	
TP6	88	67	65	65	66	62		90	84	79	78	74	76	
TP7	71	47	59	67	75	75		78	73	73	73	75	76	
TP8	100	100	100	100	100	100		100	100	100	100	100	100	
TP9	98	94	97	91	94	95		92	92	94	90	89	92	

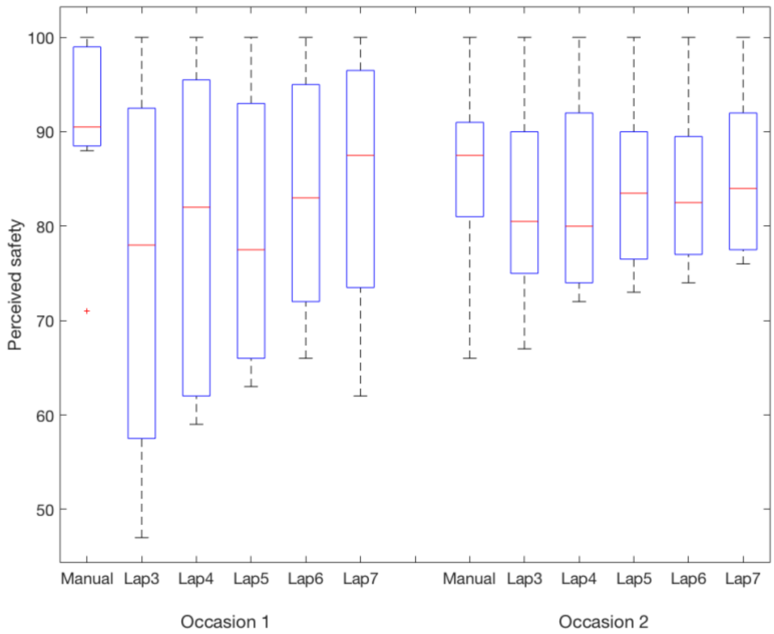


Figure 9 Boxplot showing the perceived safety self-assessment performed in the vehicle. Occasion 1 on the left and Occasion 2 on the right.

The interviews confirm that the TPs felt generally safe when traveling in the AD mode on both occasions (Table 7). The TPs reported that they had initial high level of confidence/safety towards the automated function (TP7, TP8). Some of the TPs reported an increase in safety/comfort over time as they got more comfortable with the automated function, test situation and the test track. Two of the TPs (TP8, TP9) reported that they felt similar trust as when in MD mode.

The interviews show also that there was a range of the factors contributing to the high perceived safety. Factors such as "ability to intervene", "capability of the system" and "no surprises" increased the TPs' perceived safety, while factors such as "driving style", and "road conditions" could decrease their perceived safety.

Immediately after finishing the driving session but while still being in the vehicle, the TPs were confirmed with the question "*Would you feel safe to test the AD function In real traffic*". On Occasion 1, a typical answer was "*Yes, but…"* while on Occasion 2, several of them believed they would and clearly answered "*Yes*" (Figure 10).

Table 7 The level of safety categorized into safe (i.e., TPs used the Swedish word "trygg" in their expression), and quite safe (i.e, TPs used expression such as "ganska trygg"). No stronger expression of uncertainty or unsafety were identified.

Test participant	Overall safety-Occasion 1	Progress over time during Occasion 1	Overall safety-Occasion 2	Progress over time during Occasion 2
TP2	Safe	Increase	n/a	No change
TP3	Quite safe	n/a	Safe	Increase
TP4	Safe	n/a	Safe	No change
TP5	Quite safe	No change	Safe	n/a
TP6	Quite safe	n/a	Quite safe	n/a
TP7	Safe	Increase	Safe	Decrease, increased,
TP8	Safe	No change	Safe, quite safe	No change
TP9	Safe	No change	Safe	No Change

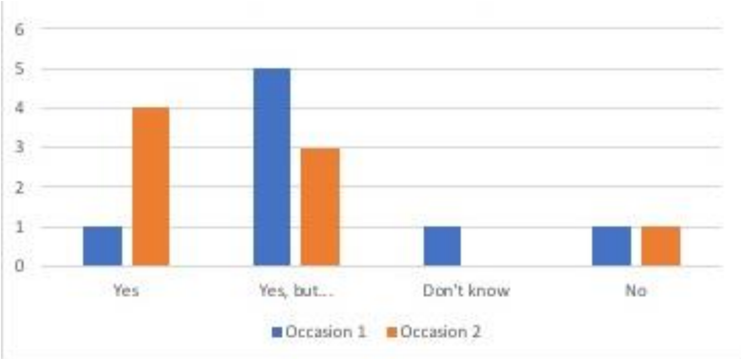


Figure 10 The test participants' answer to the post drive question "Would you feel safe to test the AD function in real traffic?"

These findings on the TPs' perceived safety are echoed by the post questionnaire that the TPs completed after Occasion 2 (Figure 6). Feeling of safety and control seem to change for the most TPs, although we still see some participants reporting there were no difference between the occasions.

3.5 Foot hovering over the brake pedal

This section relates to the assumption E: When in the AD mode and relaxed, the foot hovering over the brake pedal will decrease as compared to when in the AD mode and tensed.

Measurement of foot hovering over brake pedal was explored with the intention of assessing trust in the automated function, and readiness to take over the maneuvering control. Due to technical malfunctions, data were recorded only on Occasion 1, making comparison between the occasions difficult. During Occasion 1, data for seven TPs were recorded successfully.

Two of seven TPs showed some hovering behavior while traveling in the AD mode. An example from a TP's foot hovering is shown in Figure 11. The foot is frequently on the brake pedal in manual driving, and during a few occasions in AD mode.

The recording method as such shows some promise, but the prototype needs to be developed further to be more robust. The interpretation of data is not straight forward. For example, perceived safety could arguably be both positively (I feel safe because I am ready to brake) and negatively (I feel unsafe so I get ready to brake) correlated to brake pedal hovering. Our data is, however, too scarce to draw any definite conclusions. Further research is needed to determine how frequently this behavior occurs and what it can tell in relation to trust and perceived safety.

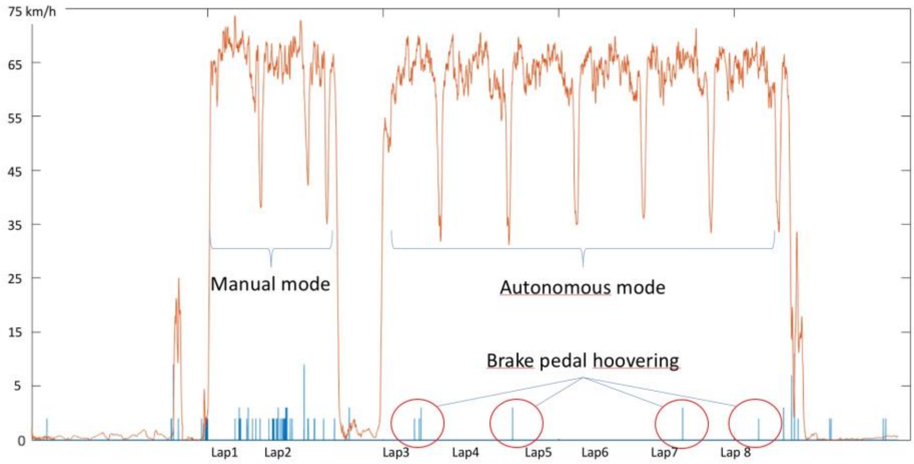


Figure 11 Brake pedal hovering during the automated mode for one TP. The diagram shows vehicle speed and instances of foot hovering the brake pedal. The dips in speed are due to the roadworks scenario on the test track.

3.6 Heart rate

This section relates to the assumption F: When in the AD mode and relaxed, the heart rate will decrease as compared to when in the AD mode and tensed.

Figure 12 shows the average heart rate for TP 2-9, with one plot for each occasion. Also, included in Figure 12 is the combined average speed of the car for Occasion 1 and Occasion 2. The average car speed shows where the different scenarios occurred. During laps 1, 2 and the beginning of lap 3 the car was driven manually by the TP. The car was in automated mode from some point in lap 3 until the end of lap 7. The big dip at the end of lap 2 and the beginning of lap 3 shows that the car was slowing down to a stand-still while the TP was instructed how to control the automated function of the car. The dips in car speed to about 40 km/h towards the end of each lap are a result of the car passing the roadwork area of the test track.

When driving manually, the TPs have on average a relatively high heart rate, especially in the beginning of Occasion 1. However, the heart rate stabilizes towards the end of manual driving session (i.e. end of Lap 2). When the TP get introduction to automation and start Lap 3, the average heart rate increases and reaches a maximum value just prior the TP gives maneuvering control to the vehicle (activate automation). After the automation is activated, the average heart rate stabilizes. However, in most cases, when there is a noticeable change in the vehicle speed due to a scenario (e.g., roadworks, overtaking), there is a short increase in the heart rate. The same pattern is observed for both occasions. That is, the TPs get a strong physiological reaction in terms of heart rate increase also on Occasion 2 just prior to activating the automation. Our interpretation is that this increase is an effect of the first encounter, and that the TP may need some more experience of using the automation to eliminate the first encounter effect.

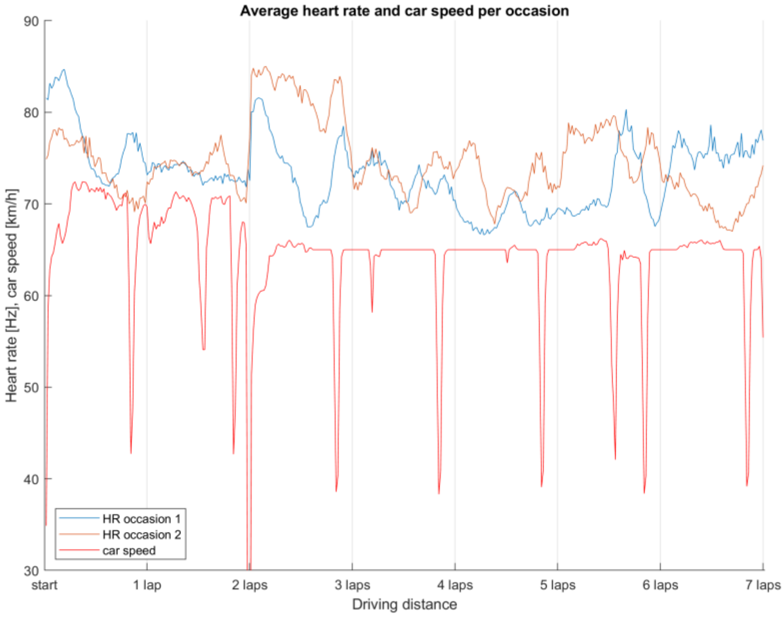


Figure 12 Average heart rate for test participant 2-9 for Occasion 1 (blue curve) and Occasion 2 (orange curve) along with the combined average speed of the experiment vehicle for Occasion 1 and Occasion 2.

3.7 Handover procedure and overall AD behavior

This section relates to the assumption F: When relaxed, the handover of control to/from vehicle will be experienced smoother as compared to when being tensed.

Overall, the handover procedure (i.e. switching between MD and AD mode) was perceived as satisfying and in line with the TPs' expectations. In the post questionnaire, the three items related to ease of use of the automated function show no difference between the Occasion 1 and 2 (Figure 6). The interviews reveal, however, that the handover experience changed from "great" on Occasion 1 to "good" on Occasion 2 (Table 8). This indicates again that the first encounter excitement has been reduced over time.

Interestingly, a few TPs assessed the vehicle behavior as more predictable, consistent and matching their expectations to a higher degree during Occasion 1 (Figure 6). This could be due to changed weather conditions between the occasions, where snow and ice on the road could have influenced how the vehicle was driving. Another possible explanation could be that the TPs were too excited during Occasion 1 and did not pay attention to such details as compared to Occasion 2 where they were more calm and relaxed.

Table 8 Self-reported characterization of the hand-over procedure

Test participant (TP)	Occasion 1- citation	Occasion 2- citation	Analysis- change over time
TP2	Very smooth	Smooth	Great → Good
TP3	Very good	It was smooth	Great → Good
TP4	Very good	Very easy to understand	Great → Good
TP5	Easy, pedagogical	Very pedagogical	Good → Great
TP6	Worked well and smoothly	It was smooth	Good → Good
TP7	Very good	Very easy	Great → Great
TP8	It felt good	It is crystal clear	Good → Good
TP9	It was smooth	It was good	Good → Good

3.8 Methodological aspects

In the post questionnaire that was completed after Occasion 2, the TPs stated that they were generally positive to the test methodology (Figure 13). They found it interesting to be involved at two re-current experiment occasions. They were also positive regarding the test equipment used. However, they suggested that the emotion assessing device could be esthetically improved to match the modern interior of the vehicle. While not clearly evident from the questionnaire, the interviews showed that the presence of the personnel (test leader, assistant, safety driver) in the backseat was something that several TPs experienced as uncomfortable and odd. As mentioned previously, this may have affected their behavior and experience of trust. As such this particular set-up may be more suitable to test aspects of user experience that do not include trust since trust is influenced by contextual factors and test situation.

From the interviews it became evident that, when experiencing the re-occurring set-up and environment, the TPs started looking for small deviations in vehicle behavior in AD mode. We also noticed that the TAM-questionnaire has a poor resolution and does not necessarily capture behavioral changes well; there were relatively small differences in the questionnaire ratings, but notable differences were highlighted in the interviews and by observing the TPs behavior in the vehicle.

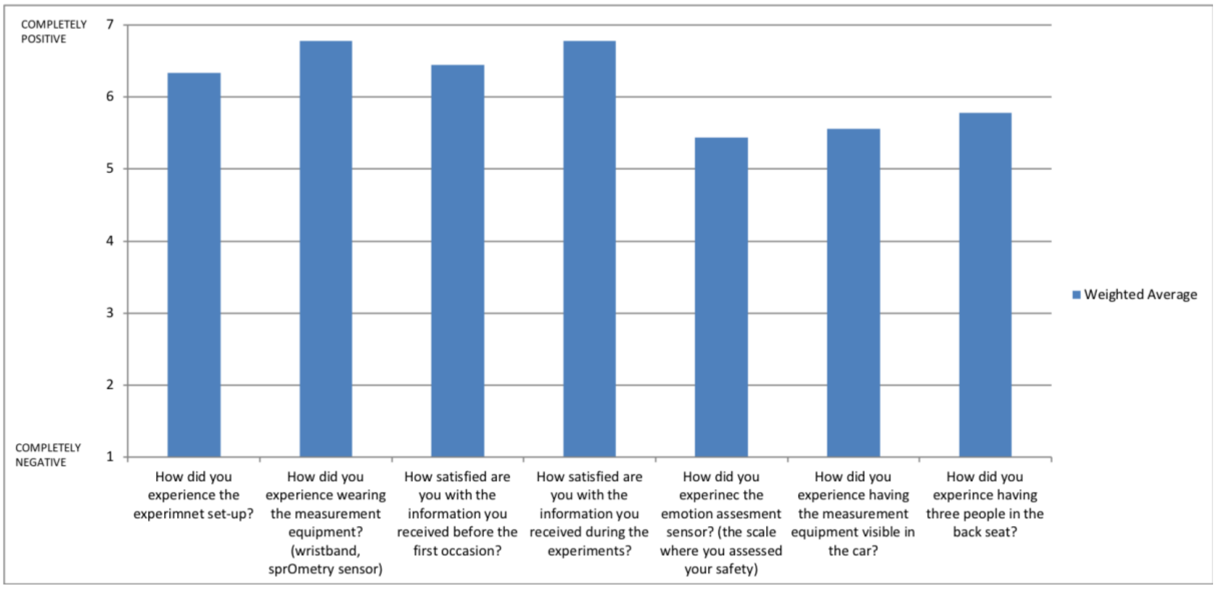


Figure 13 The test participants' opinion about the test methodology.

4 Conclusion and outlook

The preliminary analysis shows that the first encounter effects were attenuated over time. The drivers went from being exhilarated (”fun”, ”cool”, ”impressive”) on the first occasion, to a more neutral state (”positive”, ”safe”, ”good”, “relaxed”) on the second occasion. Furthermore, there were smaller variations in drivers' self-assessed perceived safety on the second occasion, and drivers were faster to engage in non-driving related activities and become relaxed (e.g., they spent more time glancing off road and could focus more on non-driving related activities such as reading). These findings suggest that exposing drivers to highly automated vehicles on two successive occasions may give more informative and realistic insights into driver behavior and experience as compared to only one occasion.

The study identified a few different quantitative factors that seem to indicate the shift from being tense to being relaxed: time to engage in non-driving related activities, glance time spent off road, heart rate and foot hovering. However, these insights are difficult to generalize due to the limited data sample. Furthermore, the study was carried out during a winter period with varying adverse weather and road conditions, which may have affected driver behavior and embedded a bias in the findings. Our recommendation is that future studies of this type should be carried out under good weather and road conditions.

The study generated valuable knowledge on pros and cons with various data collection methods, something that is valuable for future studies. In fact, these lessons-learned have already been applied in the methodology development for a similar study that we conducted recently.

5 Lessons learned, experience from testing at AstaZero

Our overall impression from the experiment at the test track AstaZero is positive. By conducting this study, we have gained knowledge on pros and cons with various data collection methods and approaches. This has enabled an easier planning of the experiments in the real-world traffic that were conducted in June 2018 within the FFI-research project *Trust in Intelligent Cars* (TIC).

One of our biggest challenges in this experiment were adverse/varying weather and road conditions. On the first experiment day, for instance, the outdoor temperature was -15°C and the roads were very slippery at the same time as it was snowing. This made roads both at the test track and in traffic difficult to access and drive on. To account for these conditions, we had to change our experimental plan, and exclude the part of the experiment in the real-world traffic, immediately before our first test participant arrived. The varying conditions created a lot of uncertainties and we never knew if it would be possible to conduct the next experiment or not. As mentioned previously, these conditions have also affected the test participants and made our study of behavioral changes over time difficult. In addition, the heating in the test vehicle needed to be high, which caused some sensor malfunctions.

The fact that test track AstaZero is rather remote made the selection and recruitment of test participants challenging, especially when considering that our test participants needed to get there at two occasions and that we were not able to offer any (economic) compensation to them. Our original plan was to get support from AstaZero in recruiting test participants, however, AstaZero recruitment pool was not operational yet. We solved this by recruiting people from RISE in Borås and enabling them to report their time on a given project. This was possible since we had only 9 test participants. However, if we would have more test participants we would need to find a different solution as it would not be feasible to offer project time to all of them. For future projects, it would be helpful to ensure that AstaZero can support recruitment and that the recruitment cost (including compensation/gift to test participants) is covered by the SAFER Open Research.

In our experiment, we needed to ensure that each test participant had the same amount of days between re-current occasions. Given that AstaZero is a popular testing facility, it was not easy to find available slots, especially when considering that this needed to be synchronized with the availability of the test vehicle. However, thanks to the helpful stuff at AstaZero, we managed to book two days at two consecutive weeks.

6 Publication and dissemination (incl. planned)

- A **workshop** has been organized in March 2018 where the preliminary results were presented to the participating stakeholders in the research project *Trust in Intelligent Cars* (Volvo Cars, Halmstad University and RISE Viktoria, ca 12 people). The workshop served as inspiration and guidance for the development of methodology for real-world experiments in the project *Trust in Intelligent Cars.*
- A **poster** summarizing the project and its results has been accepted for publication at the Automated Vehicle Symposium (AVS) 2018:
Malmsten Lundgren, V., Andersson, J., Asker, E., Habibovic, A., Klingegård, M., Lindström, D., Voronov. A., Recurrent Measurements of User Experience in AVs: A Method Development Experiment at AstaZero Proving Grounds, Poster. Automated Vehicle Symposium 2018, San Francisco, July 9-12 2018.
- A short **movie** illustrating the project has been created, and will be published at the project site after the final approval from the project partners.
- A **paper** is planned to be submitted to the Transportation Research Board (TRB) 2019, with the submission date August 1st, 2018. The preliminary title of the paper: *Studying behavioral adaptation in highly automated vehicles.*
- A request for **oral presentation** will be submitted to Transportforum 2019.
- An **oral presentation** at the SAFER Lunch Seminar is planned for fall 2018.

7 Participating partners and contact persons

RISE Viktoria AB
Lindholmspiren 3A
417 56 Göteborg
Contact: Azra Habibovic
(Senior researcher)
azra.habibovic@ri.se

Volvo Car Corporation
SE-405 31 Göteborg
Contact: Annie Rydström
(Project Manager DUX)
annie.rydstrom@volvocars.com

Smart Eye
413 27 Göteborg
Contact: Henrik Lind
(Chief Research Officer)
henrik.lind@smarteye.se