Clear and Unambiguous HMIs

Final Report

TNO Public ⟩ TNO 2025 R12106
14 November 2025

Mobility & Built Environment
www.tno.nl
+31 88 866 57 34
secretaryhelmond@tno.nl

TNO 2025 R12106 – 14 November 2025

Clear and Unambiguous HMIs

Final Report

Author(s)	J.H. (Jeroen) Hogema, M. (Marijke) van Weperen, M. (Mervyn) Fransen, H.H.S.N. (Hari) Subraveti, M.H. (Marieke) Martens, S. (Saarang) Gaggar
Internal Reviewer(s)	J.L. (Jan) Souman
Title	Clear and Unambiguous HMIs – Final Report
Number of pages	67 (excl. front and back cover)

Contents

List of Abbreviations

Abbreviation	Description
ACC	Adaptive Cruise Control
AD	Automated Driving
ADAS	Advanced Driver Assistance Systems
ADS	Automated Driving Systems
DCAS	Driver Control Assistance Systems
C&U	Clear and Unambiguous
HMI	Human Machine Interface
LCA	Lane Centring Assist
LED	Light Emitting Diode
LKA	Lane Keeping Assist

1 Introduction

As Advanced Driver Assistance Systems (ADAS) and Automated Driving Systems (ADS) evolve, driving is becoming increasingly automated—primarily through conditional automation. This leads to frequent transitions of control between the human driver and the system, requiring high-quality interaction to ensure safety and avoid mode confusion. To assess this interaction, test procedures and criteria are needed. Requirements that the system must meet are currently available only at a high and abstract level. One of the core requirement for the information a system presents to the user via the HMI (Human Machine Interface) is that it needs to be "clear and unambiguous”. These terms must be defined and operationalised, must be testable, and criteria for their assessment must be found.

The Ministry of Infrastructure and Water Management has asked TNO to further define and operationalise the term “clear and unambiguous” and also design a method suitable for finding criteria to evaluate to what extent an HMI provides “clear and unambiguous” information.

In 2023, TNO investigated the development of a quantitative measurement scale to assess Automated Driving Systems (ADS) Human-Machine Interfaces (HMIs) within the project User Centered Design of Automated Driving System Human Machine Interfaces [1] [2]. TNO developed a questionnaire to measure if the information provided to the driver by ADS (Automated Driving Systems) or DCAS (Driver Control Assistance Systems) is experienced as clear and unambiguous. This questionnaire consists of nine items, listed in Table 1.1, which have been proven to be able to differentiate between good and bad HMI designs in terms of clarity and unambiguity. It was concluded that additional research is needed to validate the questionnaire in more conditions and for different user groups.

Table 1.1: The Clear and Unambiguous Questionnaire from [1].

Item	Dutch version	English translation
1	De informatie is dubbelzinnig	The information is ambiguous
2	De informatie kan meerdere betekenissen hebben	The information can have multiple meanings
3	Ik ben onzeker over wat deze informatie kan betekenen	I am uncertain what this information might mean
4	De informatie is verwarrend	The information is confusing
5	De informatie is onduidelijk	The information is unclear
6	Ik begrijp deze informatie snel	I can quickly comprehend this information
7	De informatie is goed leesbaar	The information is legible
8	De informatie is gemakkelijk waar te nemen	The information is easy to perceive
9	De informatie wordt duidelijk gepresenteerd	The information is presented in a clear manner

The objective of the research presented here is to assess the validity and reliability of the questionnaire on realistic HMI designs and for user groups with and without experience with ADS or Driver Control Assistance Systems (DCAS). In contrast to the previous project, where the questionnaire was tested only on transitioning HMIs, it is now validated on both transitioning and static HMIs. Automation levels covered by the HMIs in this study include manual driving, L2 hands-on and L2 hands-off, where the driver must remain attentive, as well as partial driving automation (L3) where the driver can disengage but must be ready to take over. Additionally, the previous project only involved good and bad HMI designs. In the current project, it was investigated whether the questionnaire can distinguish among designs with poor, intermediate, and good quality. Furthermore, in this project, the questionnaire was tested in a driving simulator instead of a desktop environment. The simulator brings us a step closer to applying the method during actual driving.

The report is set up as follows: Chapter 2 describes the driving simulator experiment design and execution in which the Clear and Unambiguous questionnaire was applied to multiple realistic HMI designs. The stimuli, the set-up and the procedure are considered. The results of this experiment are discussed in Chapter 3. Chapter 4 presents the discussion and conclusions, as well as recommendations regarding the questionnaire.

2 Experiment Design and Execution

2.1 Introduction

In the previous study, the questionnaire has been tested in a desktop study. To bring the method closer to testing on a real system during actual driving, the questionnaire was now applied in a driving simulator. The Clear and Unambiguous questionnaire was adapted slightly (see section 2.2.4) and then used to assess to which extent the information provided by the HMI to the user was rated as clear and unambiguous. In addition, like in the previous study, questions about user responsibility were presented to gauge participants' objective understanding of the information.

2.2 Method

2.2.1 Participants

Two relevant user groups were recruited for this experiment, namely people with and without experience with level 2 automated systems. The selection for these groups was as follows (based on self-reported experience):

- Group 1: experience with L2 automated systems. People in this group had driven at least once a month with Adaptive Cruise Control (ACC) and Lane Centring Assist (LCA) for the past year.
- Group 2: no experience with L2 automated systems. People in this group had little to no experience with these systems.

Within these two experience groups, we aimed for a similar distribution of gender, age and driving experience. However, it turned out to be hard to recruit young participants without L2 experience and participants who drive often but did not have experience with L2 systems. Therefore, we needed to make some concessions. In total 38 participants participated in the experiment. Gender is shown in Table 2.1 and age and driving experience distributions are shown in Figure 2.1 and Figure 2.2. Participants were recruited by the recruitment company Link2Trials, signed an informed consent at TNO before taking part in the study and were paid for their participation. The research plan[1] was reviewed and approved by TNO's Institutional Review Board Human Research (IRB).

[1] Case number 2025-671

Table 2.1 Gender and experience with vehicle automation distribution of participants

	N	M/F/Other
Group 1: experience with L2 systems	18	14/4/0
Group 2: no experience with L2 systems	20	9/10/1
Total	38	23/14/1

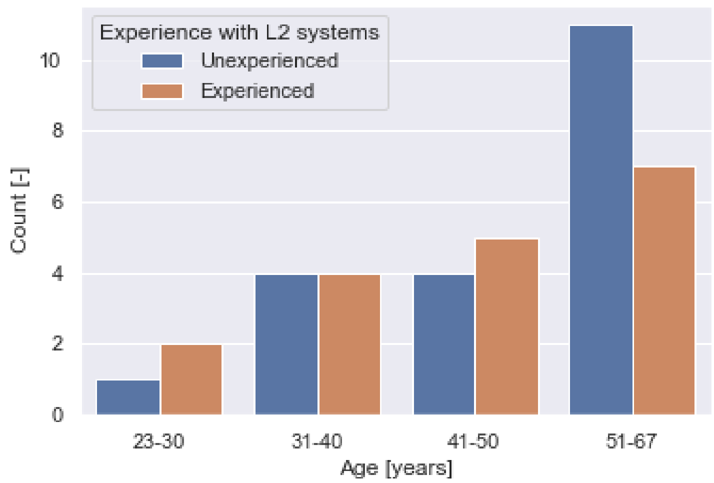


Figure 2.1 Age distribution of participants

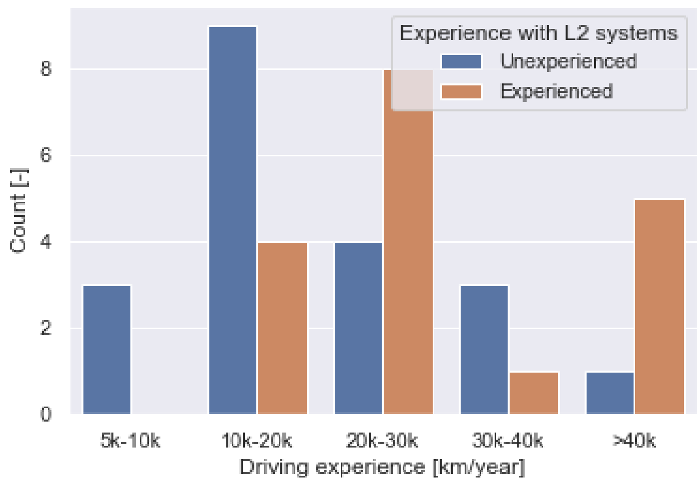


Figure 2.2 Driving experience of participants

2.2.2 Stimuli

Participants were shown 24 short movies where AD systems stayed active or were deactivated. These movies included both a virtual driving environment and an HMI, where the HMI consisted of a visual dashboard display, auditory information as well as coloured LED segments on the steering wheel. In Table 2.2 the type of HMI, the AD levels of the HMI and the quality are shown for all scenarios. Half of the scenarios consisted of a static HMI: during the entire movie, the state of the ADS and thus the content of the HMI remained the same. The other half of the scenarios contained a transition in the ADS state, from a given initial state to “Level 0”, i.e., “off”.

Table 2.2 Experiment scenarios

Scenario	Type	Version	AD levels	Intended quality in terms of clarity and ambiguity
1	Static	A	L0	Good
2	Static	A	L0	Average
3	Static	A	L0	Bad
4	Static	B	L2 Hands-On	Good
5	Static	B	L2 Hands-On	Average
6	Static	B	L2 Hands-On	Bad
7	Static	C	L2-Hands-Off	Good
8	Static	C	L2-Hands-Off	Average
9	Static	C	L2-Hands-Off	Bad
10	Static	D	L3	Good
11	Static	D	L3	Average
12	Static	D	L3	Bad
13	Transition	E	L2 Hands-on to L0	Good
14	Transition	E	L2 Hands-on to L0	Average
15	Transition	E	L2 Hands-on to L0	Bad
16	Transition	F	L3 to L0 (version 1)	Good
17	Transition	F	L3 to L0 (version 1)	Average
18	Transition	F	L3 to L0 (version 1)	Bad
19	Transition	G	L3 to L0 (version 2)	Good
20	Transition	G	L3 to L0 (version 2)	Average
21	Transition	G	L3 to L0 (version 2)	Bad
22	Transition	H	L2 Hands-off to L0	Good
23	Transition	H	L2 Hands-off to L0	Average
24	Transition	H	L2 Hands-off to L0	Bad

In the analysis, these 24 scenarios were coded in a 3x8 design, with two factors:

- “Quality”, with three levels: good, average and bad.
- “Version”, having eight levels (A-H) as listed in Table 2.2.

Besides the 24 scenarios from Table 2.2, two scenarios were created for practicing purposes at the beginning of the trial. These consisted of a static one with L3 and a dynamic one transiting from L2 to L0.

2.2.2.1 Designs of the HMIs

The HMIs designs were inspired by existing HMIs of production vehicles and by existing literature on the topic. The most important design factors that were varied are the following.

- **The amount of information elements.** HMIs vary in the amount of information they provide to the driver and this probably has an influence on how the information is perceived.
- **Differences in sizes of symbols and text.** HMIs use both smaller and bigger symbols in their designs.
- **Variety of symbols.** Car manufacturers use different symbols for their assistance systems, which could be confusing for the driver when switching brands or cars [3]. Our HMI design included different symbols for ACC, Lane Keeping Assist (LKA), Lane Centring Assist (LCA) and Level 3 automated driving. Sometimes we incuded multiple symbols for the same functionality (potentially creating confusion for 'bad' designs).
- **Various colours of symbols.** Car manufacturers use different colours to present the activation status of their ADS. Green often indicates a system is active. Red or orange inform about a warning, but the latter is sometimes also used to indicate the system is active. Blue often indicates the status of L3 functionalities, but is sometimes also shown in other symbols.
- **Positioning of symbols and text.** Given a set of information elements to be shown, their placement on the dashboard can influence to the HMI quality. The design and positioning of the symbols were varied.
- **World model.** The world model shows the driver, via the dashboard, what the ADAS or ADS has perceived in terms of other road users and the road-layout. Such a display may only show the lead vehicle or the lead vehicle in combination with the ego vehicle, but more elaborate displays show all direct surroundings. Such information may contribute to confusing the driver, suggesting higher levels of automation than is actually the case. We implemented a full world model in the driving simulator, showing all road users, including lane markings. In some of the scenarios this world model was shown to the participant.

All of the above aspects were included in the HMI design to create good, average and bad designs, according to the expert knowledge of the TNO project team. Within each of the eight versions of Table 2.2, one or more of these design aspects were varied. We did not aim to reduce the HMI quality by choosing small icons or small text size. Similar, we did not manipulate the contrast of the dashboard elements (brightness contrast or colour contrast) to vary the HMI quality. Thus, in terms of the results of the previous study [1], emphasis was on the comprehension and not so much on the perception part. The experiment also included two HMI designs from the previous study (with minor modifications, to adapt to the new dashboard design), to check if with the revised method, similar results were obtained. These were the bad and the good HMI from Version G. In that same Version G, the average HMI was the same as the good HMI. This was unintentional, but in retrospect it did allow us to test if two identical HMIs give the same results or not.

Figure 2.3 shows an example of a HMI in transition, with steering wheel LEDs activated and a world model on the dashboard. A few examples of dashboard images are shown in Figure 2.4. A detailed list of all HMI elements (dashboard, audio and steering wheel LEDs) is provided in Appendix A.

A .mp4-movie was created for every HMI, which lasted 35 seconds. Static HMIs did not change during the movie. In transition scenarios, the movie consisted of three stages: the initial phase, the transition, and the final phase. In the initial phase, a static HMI was shown. After 15 seconds, the transition phase started, in which, depending on the HMI design:

- symbols and text on the dashboard appeared, disappeared, or changed colour,

- symbols or text was flashing,
- the LEDs on the steering wheel flashed, and
- sound was presented.

Ten seconds after the start of the transition, the final phase started, in which again a static HMI was presented. At the end of the scenario, the movie was paused. Thus it was still showing the final HMI state when the participant filled in the questionnaire.

Figure 2.3 Transition phase of a scenario with flashing steering wheel LEDs.

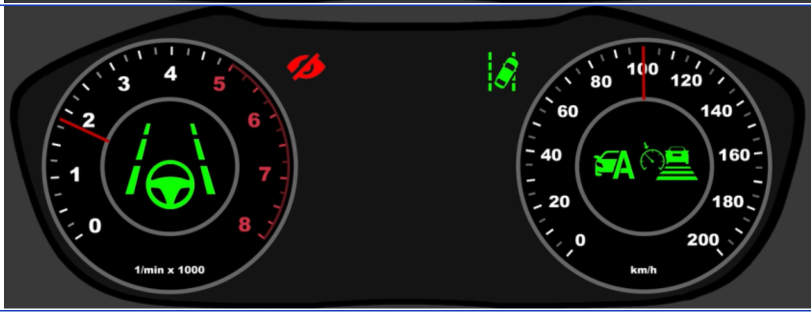
Scenario	HMI design
10	
11	
12	

Figure 2.4 Dashboard designs for scenario 10, 11 and 12.

2.2.3 Set-up

The experiment was performed in the TNO driving simulator, shown in Figure 2.5, to create a visually realistic environment to assess the HMIs. The driving simulator consisted of a BMW body on a motion platform and a 4m high and 230° cylindrical screen on which 3 projectors showed the dynamical virtual environment.

The digital cockpit of the driving simulator made it possible to show any kind of dashboard to the driver. Dimensions of the dashboard were 243 x 90 mm, with a resolution of 1920 x 720 pixels. LEDs were integrated in the rim of the steering wheel. These could show varying static or flashing colours, as shown in Figure 2.3. The audio system of the driving simulator provided the participant with sound (engine and wind), from the ego vehicle as well as from other vehicles. The motion platform from the driving simulator was stationary during the experiment.

Figure 2.5 TNO driving simulator.

During the simulation, the ego vehicle drove at a constant speed of 100km/h on a simulated three-lane motorway and did not change lanes. The lane the vehicle was positioned in (left, middle or right) was varied between scenarios, to create a little more variation in the scene (see Appendix A) . Other vehicles were present in the other two lanes and were driving slightly faster than the ego vehicle in the left lane and slightly slower than the ego vehicle in the right lane.

2.2.4 Measures

After each movie, participants answered two sets of questions. The first set consisted of the responsibility questions listed in Table 2.3. The first four questions were related to the situation at the end of the movie and what participants thought that they had to do. The fifth question checked if the participants understood, and interpreted a change in state during the movie. These multiple choice questions gave an indication if participants understood the meaning of the received information from the HMI. These questions were similar to the ones used in the previous project; the wording was slightly revised to make the statements more suited to distinguish among different automation levels. To reduce memory burden, the final state of the HMI was still visible in the car during these questions. Participants were instructed that for the transitions scenarios, these questions all referred to the final state of the scenario.

Table 2.3 Original Dutch and translated responsibility questions (where nr 5 was only for movies showing a transitioning HMI)

Item	Question in Dutch	Answer options
1	Wie moet er in deze toestand gasgeven en remmen?	"Ik doe dit helemaal zelf", "De auto, maar ik moet continu kunnen ingrijpen", "De auto doet dit zelf", "Weet ik niet"
2	Wie moet er in deze toestand sturen?	
3	Bent u verplicht om in deze toestand de omgeving in de gaten te houden tijdens het rijden?	"Ja" "Nee" "Weet ik niet"
4	Moet u in deze toestand uw handen aan het stuur houden?	
5	In welke van deze taken vond er een verandering plaats in het filmpje? Meerdere antwoorden mogelijk	"Gasgeven/Remmen", "Sturen", "Omgeving in de gaten houden", "Handen aan het stuur houden", "Weet ik niet"

Item	Question (English translation)	Answer options
1	Who needs to accelerate and brake in this state?	"I do this all by myself", "The car, but I constantly need to be able to intervene" "The car does this all by itself" "I don't know"
2	Who needs to steer in this state?	
3	Are you obligated to keep an eye on the environment during driving in this state?	"Yes" "No" "I don't know".
4	Are you obligated to keep your hands on the steering wheel in this state?	
5	Which of the following tasks changed during the transition in this movie? Multiple answers possible	"Accelerating/Braking", "Steering", "Keeping an eye on the environment", "Keeping hands on the steering wheel" and/or "I don't know"

Next, the participants filled out the Clear and Unambiguous questionnaire (see Table 2.4) for the HMI they had experienced, answering all 9 questions on a 7-point scale. The questionnaire has been slightly adapted from the one in the previous project, by inverting item 3, 5, 7 and 8. This was preferred since statements belonging to dimension 1 were all negative and dimension 2 all positively formulated, which was considered undesirable. With the new formulation, we stayed close to the original wording, while maintaining a simple word choice. Further, to keep the formulation simple, we realised the inversion by other means than inserting the word 'not'.

Table 2.4 Clear and Unambiguous questionnaire.

Item	Question in Dutch	English translation
1	De informatie is dubbelzinnig	The information is ambiguous
2	De informatie kan meerdere betekenissen hebben	This information can have multiple meanings
3	Ik weet zeker wat deze informatie betekent	I am certain what this information means
4	De informatie is verwarrend	This information is confusing
5	De informatie is duidelijk	This information is clear
6	Ik begrijp deze informatie snel	I can quickly comprehend this information
7	De informatie is slecht leesbaar	The information is illegible
8	De informatie is moeilijk waar te nemen	The information is hard to perceive
9	De informatie wordt duidelijk gepresenteerd	The information is presented in a clear manner

The labels on the Likert scale were as shown in Table 2.5.

Table 2.5 Likert scale labels.

Nr	Label in Dutch	English translation
1	Sterk mee oneens	Strongly Disagree
2	Oneens	Disagree
3	Enigszins oneens	Somewhat Disagree
4	Niet eens / niet oneens	Neither agree nor disagree
5	Enigszins mee eens	Somewhat Agree
6	Eens	Agree
7	Sterk mee eens	Strongly Agree

After all questionnaires were completed, the participant answered some final questions during a semi-structured interview, as listed in Table 2.6. These questions were intended to provide more insight into the reasoning behind the answers to the Clear and Unambiguous questionnaire. To minimise the differences in interviewing technique between test leaders, test leaders were only allowed to ask a few clarifying questions in addition to the questions above.

Table 2.6 Open-ended post-experiment questions

Nr	Question (in Dutch)	English translation
1	"Waren er bepaalde dingen die de informatie duidelijk maakten?"	"Were there specific things that made the information clear?"
2	"Waren er bepaalde dingen die de informatie verwarrend maakten?"	"Were there specific things that made the information confusing?"
3	"Waren er bepaalde dingen die u een gevoel van onzekerheid gaven?"	"Were there specific things that gave you a feeling of uncertainty?"
4	"Waren er bepaalde dingen die u een gevoel van zekerheid gaven?"	"Were there specific things that gave you a feeling of certainty?"
5	"Heeft u bepaalde dingen gemist?"	"Did you miss specific things?"

2.2.5 Procedure

The procedure is visualised in Figure 2.6. Each participant started with an introduction of the experiment and signed the informed consent. The participant was informed that he/she would experience the movie in the driving simulator, but would not have to steer or brake themselves. When the movie included a transition, the movie was played twice before the questionnaire appeared. Some movies included a voice-over saying "Je besluit het stuur vast te pakken" ("You decide to grab the steering wheel"), which indicated what the driver was doing in that situation, but participants did not have to perform the action themselves. An example was played for the participants before they stepped into the driving simulator. Thereafter, they filled in a questionnaire regarding their gender, age (category), driving experience (category), the number of years in possession of a drivers license (category) and their experience with ACC and Lane Centring. This part of the procedure took about 15 minutes.

Participants always started with one static and one transition practice movie (scenarios 25 and 26). After that, all 24 scenarios from Table 2.2 were presented in a random order. The transition movies were played twice before the participant filled in the questionnaire. This was done to give participants sufficient opportunity to observe well what happened during the scenario. The static movies, containing less information to observe, were shown only once before participants filled in the questionnaires.
Questionnaires were filled in in the vehicle, on a tablet that was mounted on the centre console, right next to the participant. After half of the scenarios were completed, the participant had a break for about 10 minutes before the second half of the scenarios were administered. Overall, this part of the procedure took about 1.5 hours.

After the participant had completed all scenarios, they performed the semi-structured interview, filled in a final form regarding comments or complaints about the study and completed a payment form. The complete procedure took about 2 hours.

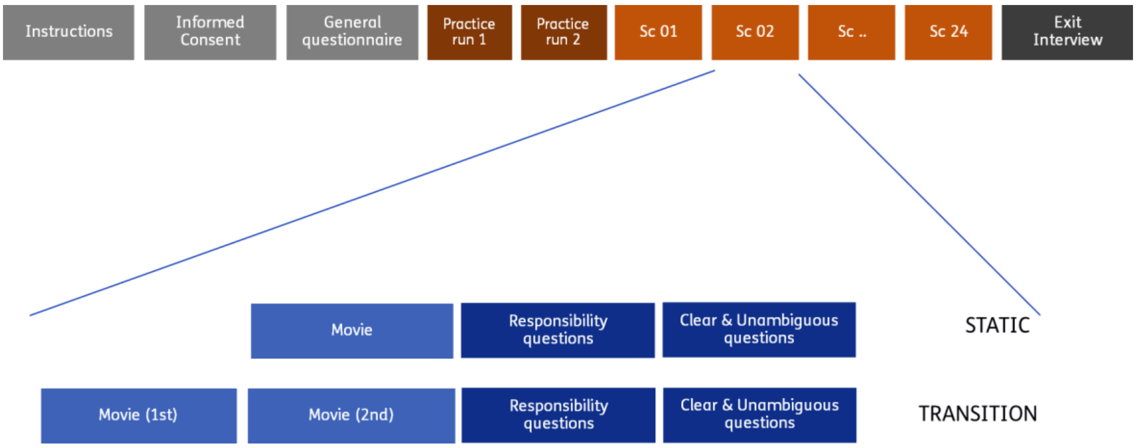


Figure 2.6 Experimental procedure.

2.2.6 Analysis

To analyse the data, a similar approach as in the 2023 project was used. For the Clear and Unambiguous questions, initially the Likert scores for individual questions were analysed, broken down by scenario quality level. Next, factor analysis was applied to investigate to what extent the set of C&U questions showed the same underlying factors. Initially, the raw

Likert scale scores on the questions from Table 2.4 were scored as a value in the range of 1 to 7. Before starting the factor analysis, these raw scores were re-coded to a score ranging from -3 to +3. Furthermore, the scale was inverted if needed, such that positive scores always matched a favourable assessment of the information. The factor loadings from the factor analysis were used to obtain values for underlying factors, also on a scale from -3 to +3.

Next, Linear Mixed-effects Model (LMM) analysis was applied on the resulting factors. Factors in the analysis were the HMI quality level, the scenario version, the interaction between these two, the experience with L2 systems, and the interaction between HMI quality and L2 experience. Finally, participant ID was included as a random factor, allowing for different intercepts per participant. To test for differences among the various levels, a type III ANOVA was performed on the LMM. The a-priori expectation was that there would be differences among the three HMI quality levels, expecting higher scores as the HMI quality level increased. Therefore, planned comparisons were applied to test for significance of these differences. For other factors, Tukey post-hoc tests were conducted. Software and packages used in these analyses are listed in Appendix C.

To analyse the responsibility questionnaire results, the responses were compared against the correct answers (given the level of automation of the ADS, and the resulting responsibilities of the driver and the system). These binary scores were processed into percentages of correct answers as a function of the HMI quality. These percentages were subjected to tests for statistical significance; see Appendix C for details.

Next, it was investigated is if there was a relationship between the Comprehension scores and Perception scores on the one hand, and the correctness on the responsibility questions on the other. Intuitively, one might expect that better HMIs will have higher percentages of correct answers, but this needs to be assessed. After all, it would be concerning if users think they have a good comprehension, but their understanding is incorrect. To investigate this, the average Perception and Comprehension scores were determined for each scenario, as well as the percentage of correct responses for each of the four responsibility questions. Linear regression models were fitted to investigate relationships between the responsibility question on the one hand, and the Perception and Comprehension scores on the other.

For analysing the open-ended questions, codes and themes were derived from the responses through a thematic analysis (Braun and Clarke, 2006). A thematic analysis is a qualitative research method used to identify, analyse, and interpret patterns (called themes) within a set of data, such as interview transcripts, open-ended survey responses, or written texts. In thematic analysis, *codes* are short labels that identify specific pieces of data, while *themes* are broader patterns that emerge from grouping related codes together. Codes describe what is in the data, whereas themes interpret and explain the underlying meaning across the dataset. One project team member reviewed the content of the responses to extract these codes and themes. This was done separately for each question.

3 Results

3.1 Participant characteristics

Correlations among the participant characteristics are shown in Figure 3.1.

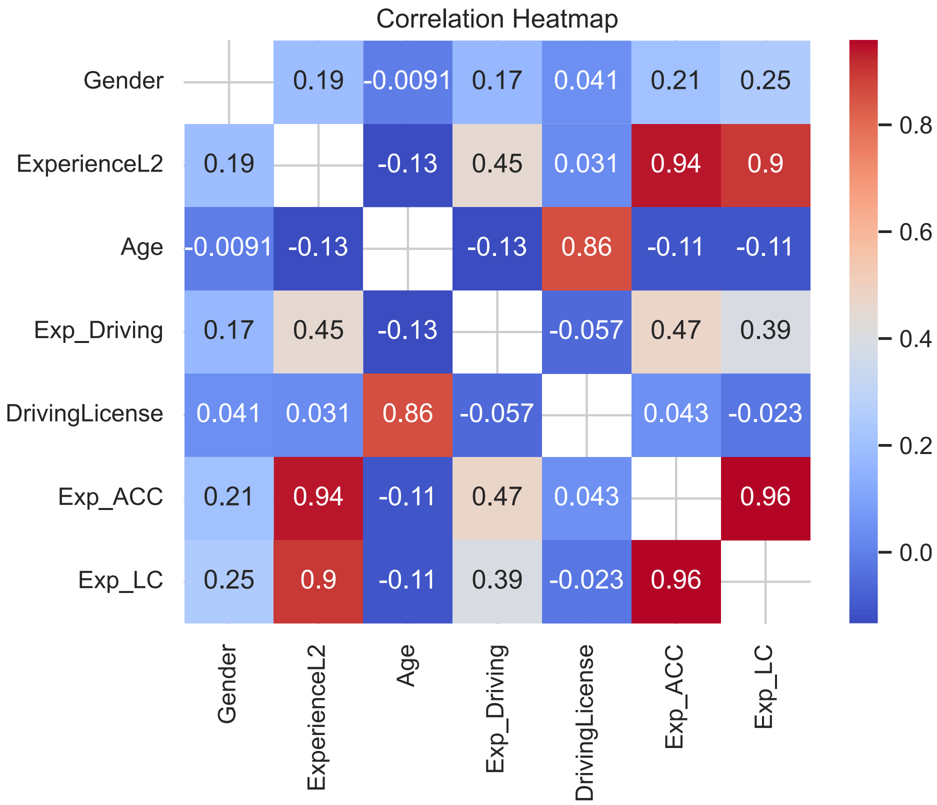


Figure 3.1 Correlations among the participant characteristics.

There was a strong correlation between participants' experience with ACC, lane centring systems, and Level 2 automated systems. All of these correlations were statistically highly significant [all $p<0.001$]. Furthermore, there also was a strong correlation between age and driving experience in years [$p<0.001$]. Otherwise, there was no significant correlation between gender and any other characteristic [all $p>0.44$]. Finally, there was a moderate (though statistically significant) positive correlation between driving experience on the one hand, and experience with L2 / ACC / LC systems on the other [all $p<0.02$].

3.2 C&U Questionnaire

3.2.1 Likert scales

Cronbach's alpha for the Clear and Unambiguous Questionnaire was 0.92 for all items together, indicating high internal consistency.

Figure 3.2 shows the participant responses per Clear and Unambiguous Questionnaire item, separated by the quality level but otherwise collapsed across all scenarios. For most questions, the graphs show that participants judged the good HMIs more positively and the bad HMIs more negatively. The only exceptions are questions 7 and 8. On these items (difficult to read, difficult to perceive), more than 90% of participants disagreed or disagreed strongly, irrespective of the quality level of the HMI.

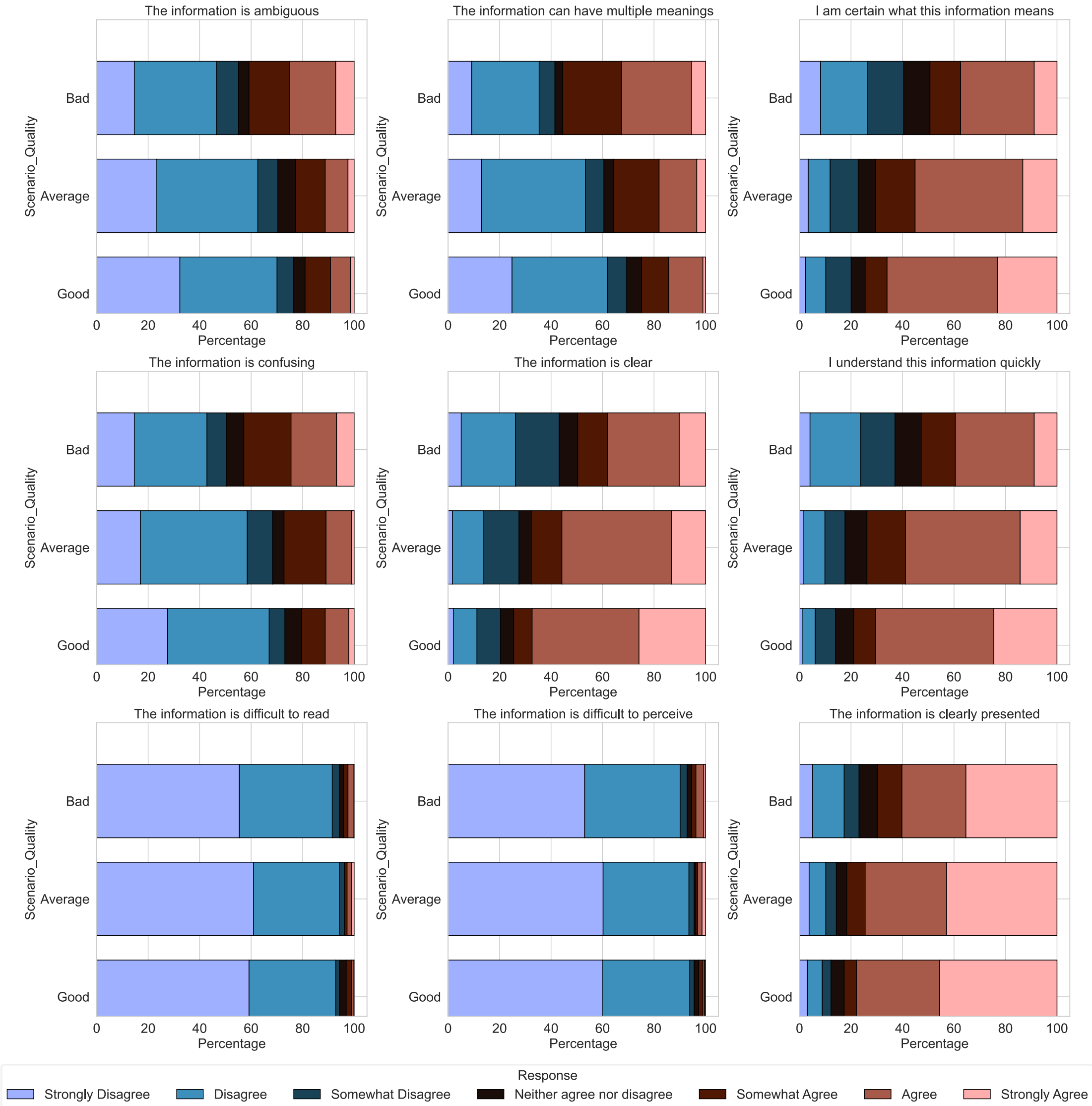


Figure 3.2 Clear and Unambiguous Questionnaire responses on original scales.

3.2.2 Factor analysis

A factor analysis was performed keeping factors with eigenvalues above 1. As in [1], oblique rotation was used as the underlying factors may be correlated. Two factors were extracted that together explained 79.4% of the variance. The scree plot and a low eigenvalue of the third factor (0.48) justified the choice to not include more factors.

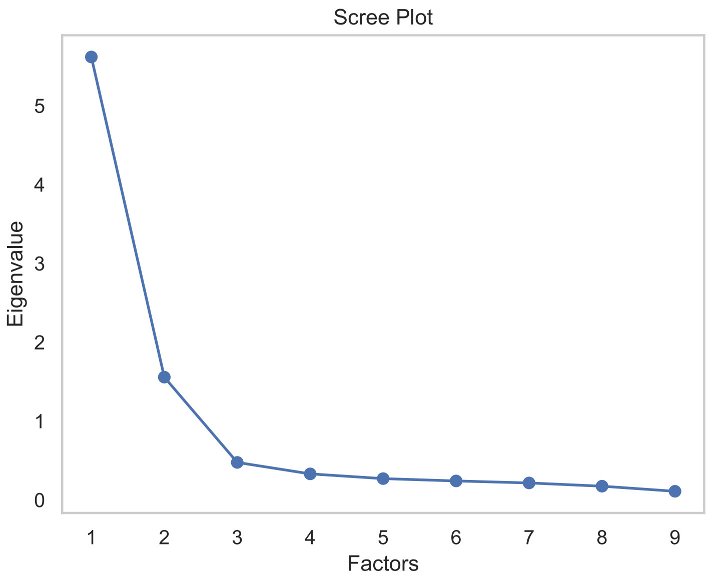


Figure 3.3 Scree plot of the factor analysis.

Table 3.1 Pattern Matrix with factor loadings when extracting two factors ranked highest to lowest (values <0.3 are suppressed; * are negatively phrased questions that were inverted before the factor analysis).

Item	Factor 1	Factor 2
I am certain what this information means [Q3]	0.93	
This information is clear [Q5]	0.93	
This information is confusing [Q4] *	0.91	
The information is ambiguous [Q1] *	0.87	
This information can have multiple meanings [Q2] *	0.86	
I can quickly comprehend this information [Q6]	0.86	
The information is presented in a clear manner [Q9]	0.51	0.31
The information is hard to perceive [Q8] *		0.91
The information is illegible [Q7] *		0.87

In line with the 2023 results [1], we labelled the factors as *Comprehension* (Factor 1) and *Perception* (Factor 2). Comprehension and Perception scores were obtained by computing weighted averages of the item responses, using the factor loadings from Table 3.1 as the weights. The table also shows that one question contributed to both factors: Q9, “The information is presented in a clear manner”. The Cronbach’s alphas were 0.95 and 0.72 for the items in the comprehension factor and in the perception factor, respectively. Removal of Q9 yielded Cronbach's alpha of 0.95 and 0.86, respectively. This shows that Q9 degraded the Perception factor rather than strengthening it. Therefore, in the calculation of the Perception scores, only Q8 and Q7 were used. The resulting factors are analysed in the next section.

3.2.3 Comprehension and Perception scores

Distributions of the Comprehension and Perception scores as obtained with the results from the factor analysis are shown in Figure 3.4 and Figure 3.5. Since HMI quality was designed to vary mainly in Comprehension, not in Perception, larger variation was expected for the Comprehension scores compared to the Perception scores. This was indeed the case. The Comprehension score covered the entire range from -3 to +3, with a peak around +2. The Perception score was much more concentrated: more than 90% of the observations are in the range between 2 and 3.

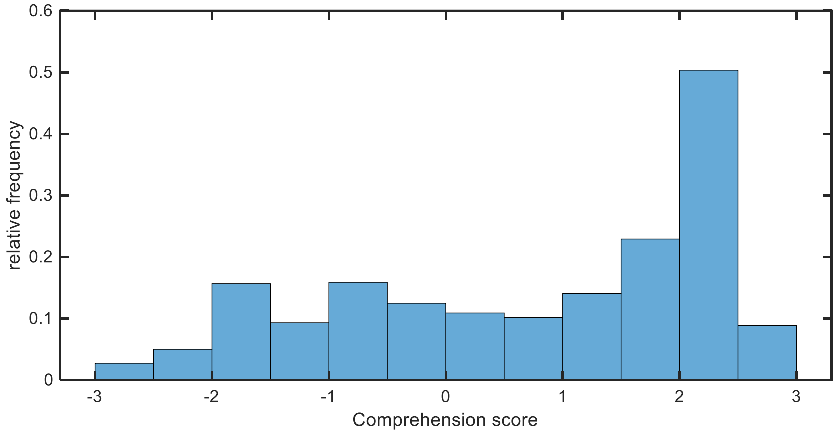


Figure 3.4: Distribution of the Comprehension score.

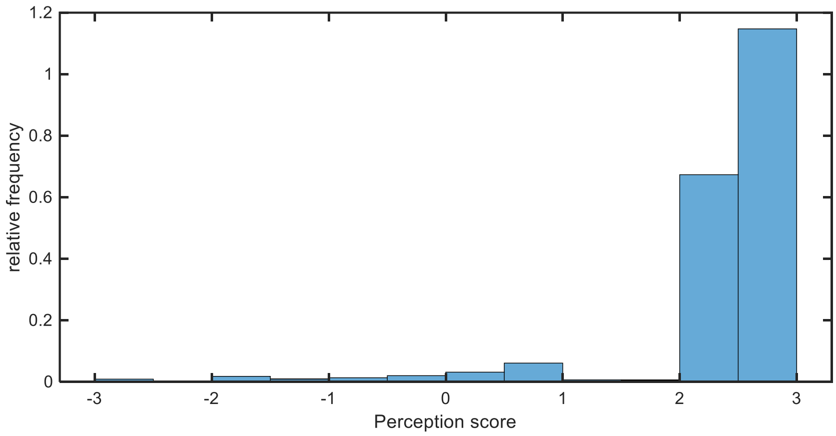


Figure 3.5: Distribution of the Perception score.

The Comprehension scores and Perception score were analysed in Linear Mixed Models (LMMs).The ANOVA results for the Comprehension score are given in Table 3.2. The effects of HMI quality, version, as well as their interaction were statistically significant [all p<0.001].

Table 3.2: ANOVA results for the Comprehension score.

	Sum Sq	Mean Sq	NumDF	F	p
Quality	174.27	87.14	2	61.68	<0.001
Version	68.49	9.79	7	6.93	<0.001
ExperienceL2	0.66	0.66	1	0.47	0.500
Quality:Version	157.39	11.24	14	7.96	<0.001
Quality:ExperienceL2	1.69	0.84	2	0.60	0.550

The marginal means of the Comprehension score are shown in Figure 3.6 as a function of HMI quality. Planned comparisons showed that the differences among all three levels were statistically significant [all p<0.001].

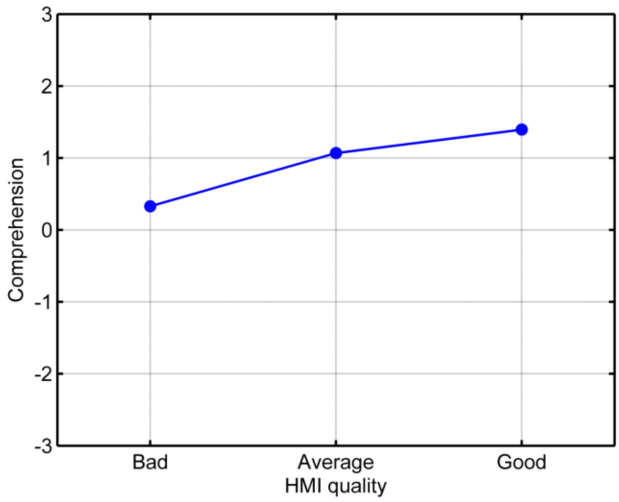


Figure 3.6: Average Comprehension score as a function of HMI quality.

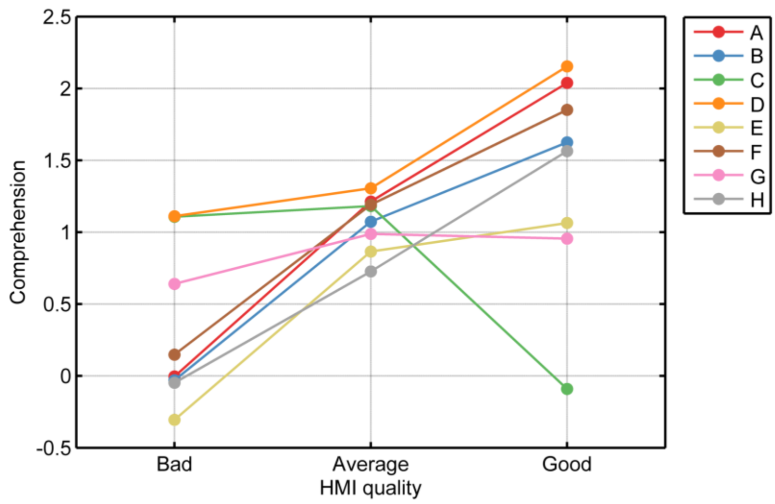


Figure 3.7: Comprehension score as a function of HMI quality and scenario version (A-H).

The interaction effect between HMI quality and scenario version is shown in Figure 3.7. The results from the paired comparisons per scenario version are shown in Table 3.3.

The general pattern was that the Comprehension score increased with the HMI quality level, in line with the main effect. In six out of eight versions, the scores for good HMIs were significantly higher than for bad HMIs. Of these six versions, four versions had a significant difference between bad and average, as well as between average and good (A, B, F, H), whereas two versions had only one of these differences statistically significant (D, E). The remaining two out of eight versions showed no significant effect (G) or a reversed effect (C). As explained in Section 2.2.2.1, the average and good scenarios of Version G were, in fact identical. Therefore, it was to be expected that there was no difference between these conditions [p=0.88].

Table 3.3: Comprehension scores: P values in planned comparisons among the HMI quality levels, for scenario versions A-H.

Version	Bad-Average	Average-Good	Bad-Good
A	**<0.001**	**0.003**	**<0.001**
B	**0.001**	**0.050**	**<0.001**
C	0.774	**<0.001**	**<0.001**
D	0.477	**0.003**	**<0.001**
E	**<0.001**	0.491	**<0.001**
F	**<0.001**	**0.019**	**<0.001**
G	0.205	0.881	0.264
H	**0.005**	**0.003**	**<0.001**

The ANOVA results for the Perception score are given in Table 3.4. Here, the quality level was the only significant factor [p=0.03]. Furthermore, there were marginally significant interaction effects between quality and version [p=0.06] and between quality and experience with Level 2 systems [p=0.09]. The marginal means of the quality main effect are shown in Figure 3.8. Planned comparisons revealed that the difference between the bad and both other levels was statistically significant [both p<0.04], whereas no significant difference was found between average and good [p=0.85].

Table 3.4: ANOVA results for the Perception score.

	Sum Sq	Mean Sq	NumDF	F	p
Quality	3.85	1.92	2	3.59	0.02813
Version	4.80	0.69	7	1.28	0.25749
ExperienceL2	0.73	0.73	1	1.36	0.25116
Quality:Version	12.40	0.89	14	1.65	0.06091
Quality:ExperienceL2	2.61	1.30	2	2.43	0.08837

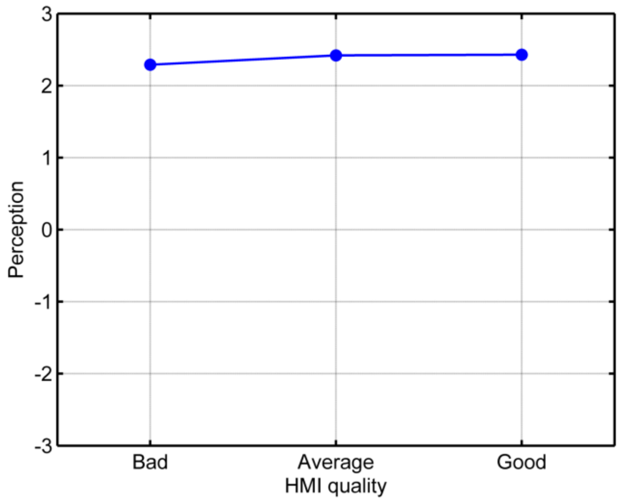


Figure 3.8: Average Perception score as a function of HMI quality.

The interaction effect between HMI quality and scenario version is shown in Figure 3.9. Planned comparison results for differences among the three HMI quality levels are shown in Table 3.5 for the eight scenario versions.

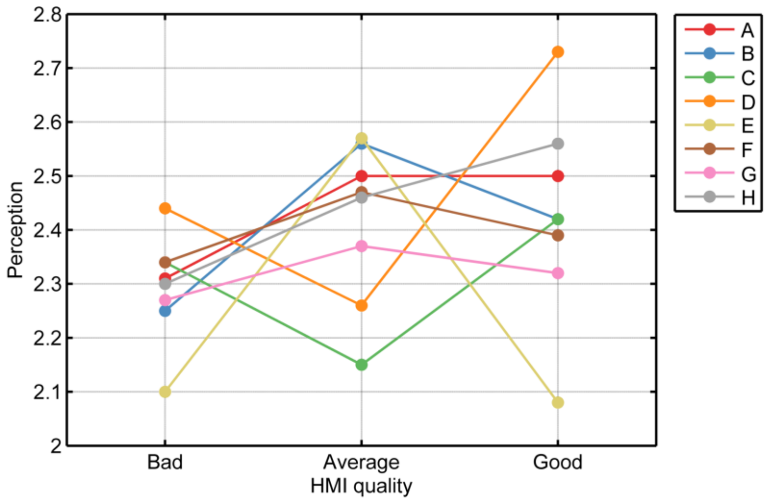


Figure 3.9: Perception score as a function of HMI quality and scenario version (A-H).

In five out of eight versions, no significant differences were found among any of the quality levels (version A, C, F, G, H). As explained in Section 2.2.2.1, the average and good scenarios of Version G were in fact identical. Therefore, it was to be expected that there was no difference between these conditions [p=0.75]. In two out of eight versions, there was a statistically significant difference between two of the three quality levels, suggesting at least *some* effect of the HMI quality level on the Perception score (B, D), in the expected direction (i.e., an increasing Perception score when the HMI quality improves). In one of the eight versions (E), the perception score was highest for the average HMI, significantly higher than either the bad or the good HMI.

Table 3.5: Perception scores: P values in planned comparisons among the HMI quality levels, for scenario versions A-H.

Version	Bad-Average	Average-Good	Bad-Good
A	0.263	0.996	0.269
B	**0.069**	0.439	0.296
C	0.271	0.113	0.623
D	0.293	**0.006**	**0.082**
E	**0.005**	**0.004**	0.947
F	0.428	0.640	0.743
G	0.552	0.754	0.777
H	0.338	0.578	0.130

As Table 3.4 shows there was a marginally significant interaction effect between HMI quality level and L2 experience [p=0.09]. This interaction effect is shown in Figure 3.10. In Tukey tests, no significant differences were found within the with “no experience” [all p>0.6]. In the group “with experience", the difference between good and bad was statistically significant [p=0.005] and the difference between average and bad was marginally significant [p=0.09].

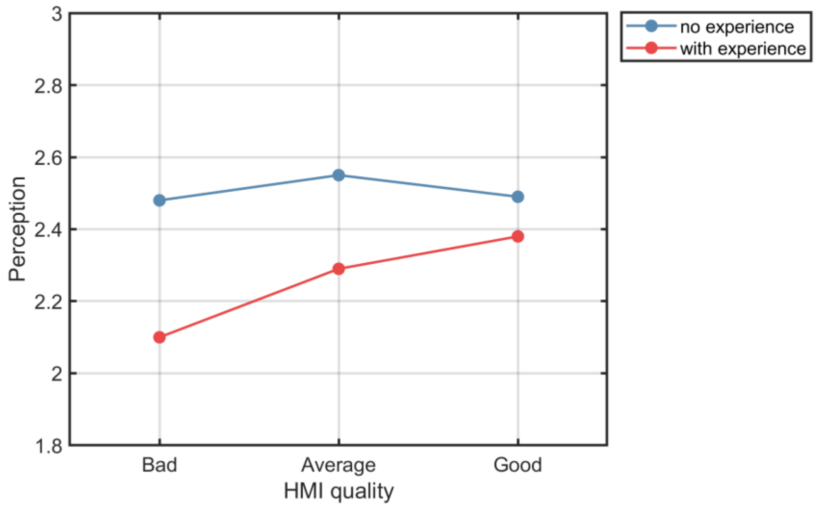


Figure 3.10: Perception score as a function of HMI quality and Level 2 experience

In summary, most Perception scores were above +2, suggesting that most HMIs were easy to perceive, regardless of the HMI quality level. While there was a significant effect of HMI quality—bad HMIs scored lower than average or good ones—most scenario versions showed no significant differences. Only a few scenarios showed expected (2x) or unexpected (1x) effects. While no main effect of L2 experience on the Perception score was found, results indicated that experienced participants did show an effect of HMI quality on the Perception score, whereas inexperienced participants did not.

3.3 Responsibility questions

For the four responsibility questions from Table 2.3, the number of correct/incorrect responses were assessed per HMI quality level. The *"I don't know"* answers were scored as incorrect. The percentages of correct answers are shown in Figure 3.11. For each question, a statistical test was performed to assess whether the three quality levels yielded the same proportion of correct responses or not. The results are summarized in Table 3.6.

Table 3.6: Statistical test results for the responsibility questions: main effect of HMI quality level

Question	X^2	df	p
accelerate-brake	23.8	2	<0.001
steering	11.1	2	0.003
awareness	7.3	2	0.025
hands on SW	12.7	2	0.002

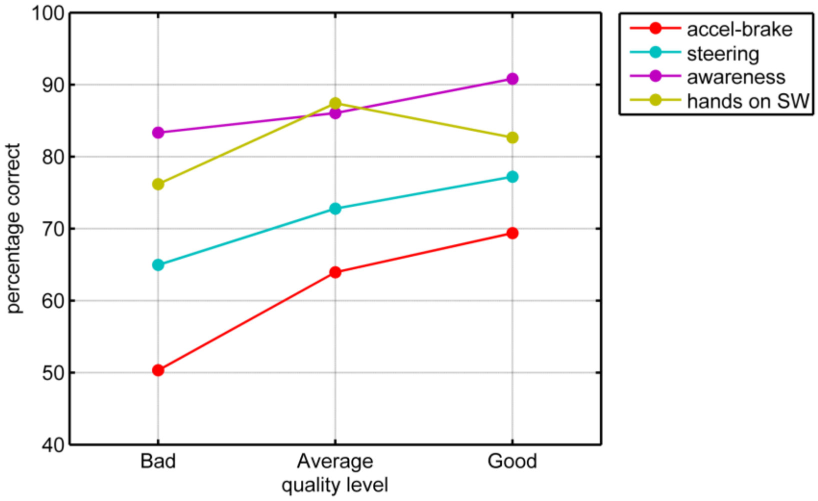


Figure 3.11: Percentage correct for the responsibility questions as a function of HMI quality level

The overall tendency was that the percentage of correct answers increased as a function of the HMI quality. Only in the "hands on steering wheel" question, the highest percentage of correct answers was obtained in the Average HMI quality group. Pairwise comparisons with Holm correction were carried out to assess significance of the differences among the three quality levels. As Table 3.7 shows, the difference between *good and bad* HMIs was statistically significant for the first three questions, and not significant for the "hands on wheel" question. The difference between *average and bad* HMIs was statistically significant for the "accelerate/brake" and for the "hands on wheel" question, and marginally significant for the "steering" question. Comparing the good and average HMI's, no significant differences were found in any of the four questions. Overall, for all responsibility questions, as the quality level increased, the percentage of correct answers either increased or did not change significantly.

Table 3.7: P-levels in pairwise comparisons with holm correction of the percentage correct in the responsibility questions.

Responsibility question	Good-Average	Average-Bad	Good-Bad
1 Accelerate/brake	0.189	0.002	<0.001
2 Steering	0.253	0.100	0.004
3 Awareness	0.187	0.423	0.030
4 Hands on wheel	0.133	0.002	0.133

The next step was to investigate if the percentage of correct answers depended on the Comprehension score or on the Perception score. The results from the regression analysis are shown in Figure 3.12 and Figure 3.13. The regression coefficients and their statistically significance are presented in Table 3.8.

For the question on accelerating/braking, a significant correlation was found with the Comprehension score and a marginally significant correlation with the Perception score. These correlations show a higher percentage correct for a higher Perception or Comprehension score. Otherwise, no significant correlations were found.

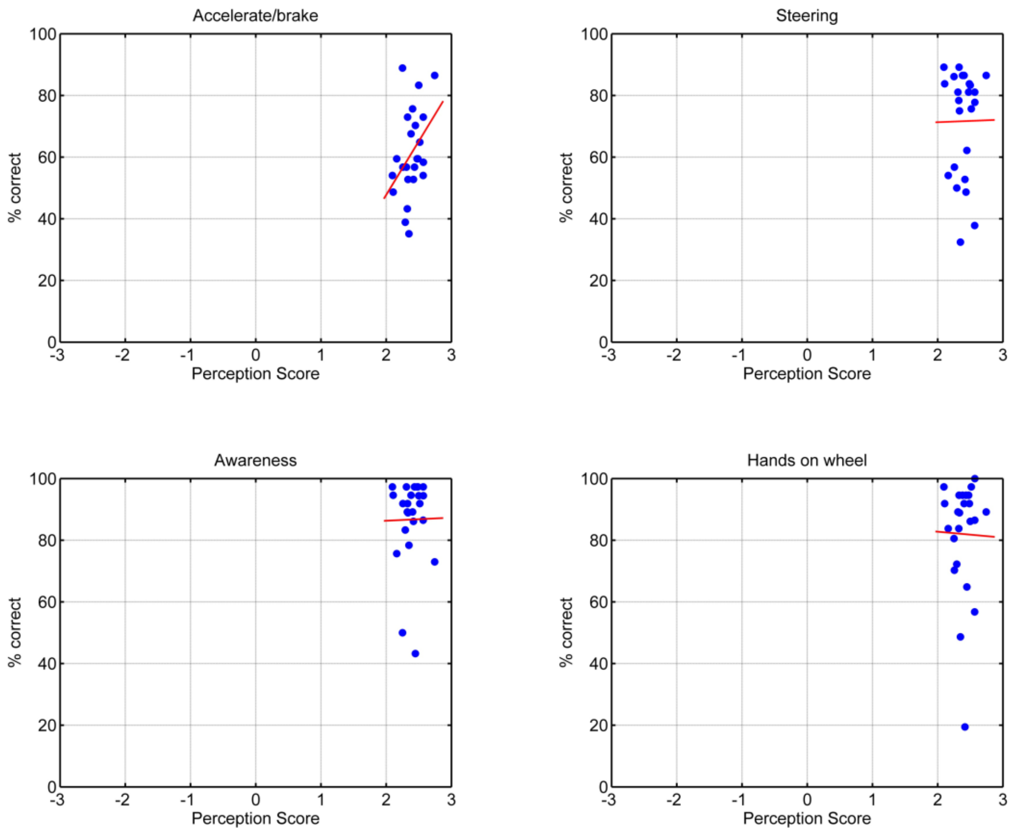


Figure 3.12: Responsibility questions: percentage of correct responses as a function of the Perception score.

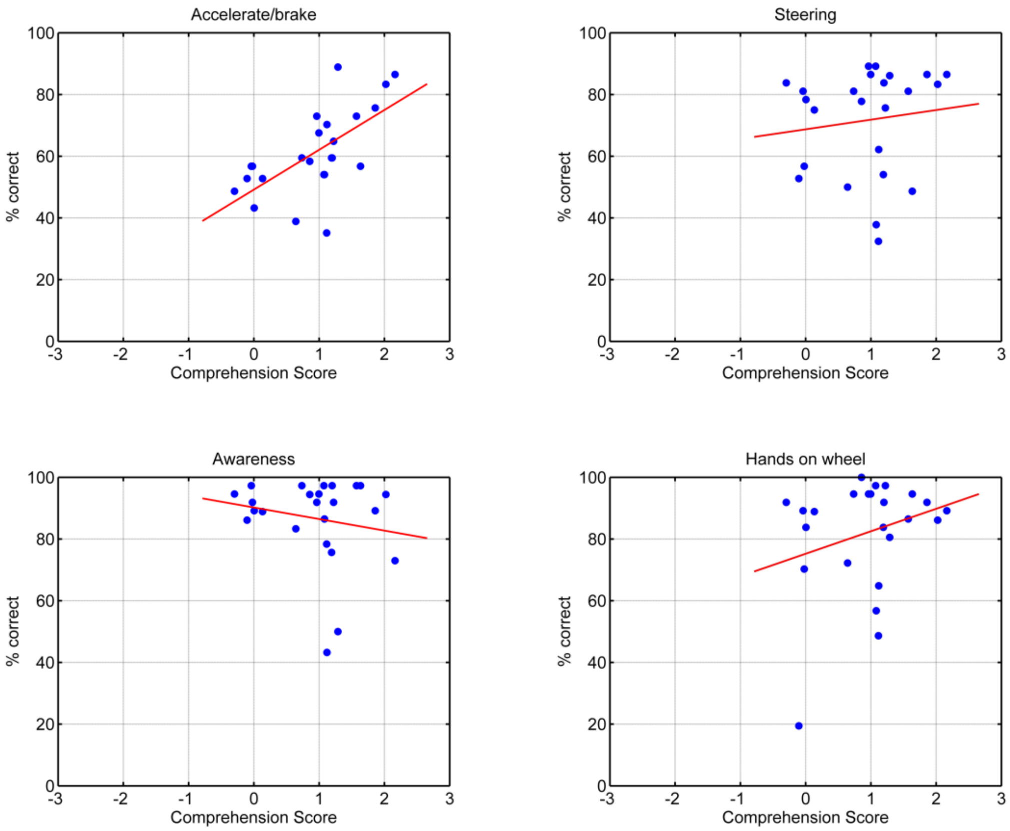


Figure 3.13: Responsibility questions: percentage of correct responses as a function of the Comprehension score.

Table 3.8: Regression results for responsibility questions and Clear and Unambiguous factors.

Responsibility question	Factor	regression coefficient	p
1 Accelerate/brake	Perception Score	34.84	0.061
1 Accelerate/brake	Comprehension Score	12.90	0.001
2 Steering	Perception Score	0.84	0.972
2 Steering	Comprehension Score	3.12	0.563
3 Awareness	Perception Score	1.01	0.959
3 Awareness	Comprehension Score	-3.74	0.394
4 Hands on wheel	Perception Score	-1.90	0.942
4 Hands on wheel	Comprehension Score	7.29	0.205

3.4 Exit interview

The verbal responses of the participants to the open-ended questions were written down by the experimenter in real-time. In the analysis phase, a thematic analysis was conducted, as explained in Section 2.2.6. The codes and themes that were derived are presented below per question.

Aspects that made the information clear
When it comes to clear information, a general theme emerged: “communication of information”. Within this theme, “text” was mentioned by 25 participants, “colours” by 14 participants, sometimes specifically mentioning green (3) and red (3). Furthermore 8 participants mentioned “auditory cues”, followed by “simplicity”, “blinking indications”, “LEDs in the steering wheel” and “displaying the cars and environment”, all being mentioned twice.

Aspects that made the information confusing
3 of the 38 participants indicated that they did not experience any form of confusion during the study. When it comes to unclear information, two general themes emerged: “communication of information” and “dashboard design”. Within the theme of dashboard design, 21 participants mentioned “unclear icons” and 19 mentioned “colour”. Within the theme of information 9 participants mentioned “too much information”, 5 mentioned “double information”, 3 mentioned “unclear information”, 2 mentioned “incorrect information” and 1 mentioned "too little information”.

Aspects that gave you a feeling of uncertainty
First of all, 11 of the 38 participants indicated that they did not experience any uncertainty. The participants that did experience uncertainty were broad in explaining what they felt uncertain about. This led to a vast amount of codes and the largest amount of themes. The themes are: “information”, “unclarities”, “colours”, “system errors" and "trust”. For the theme information 5 participants mentioned “double information”, 4 mentioned “no explanation”, also 4 mentioned “unknown information”, also 4 mentions of “contradicting information”, 3 mentioned "too much information” and 2 mentioned “too little information”. For the theme unclarities the participants mentioned “unclarity” 6 times, followed by 4 mentions of “unclear icons”, and mention of each of the following: “no confirmation”, “no consistency” and “confusing”. Colour was mentioned twice as general “colour” and twice as “yellow”, followed by “orange” and “red” that were both mentioned once. The theme system error with one code: “system error”, was mentioned 3 times. Lastly the theme trust was made up of two codes that were both mentioned once: “letting go of trust” and “trusting driving style of the autonomous car”.

Aspects that gave you a feeling of certainty
Out of the 38 participants, 2 indicated that they did not experience any certainty. The derived themes for certainty are: “information”, “safety”, “control” and “dashboard design”. Information was the broadest theme, here 13 participants mentioned “clarity”, just like “text” that was also mentioned 13 times. Next to that “experience” was mentioned 6 times, followed by 3 mentions of “not too much information”. Lastly “simplicity” and “easy to understand” where both mentioned once. For the theme control 3 people mentioned “being in control” was linked to their certainty. In dashboard design the colour “green” was mentioned twice, followed by “icons” and “display of car and surroundings” being mentioned once. Lastly, the theme safety got mentioned twice as a mention of “safety”.

Aspects that you missed
16 of the 38 participants indicated that they did not experience the feeling of missing anything during the experiment. The two themes that arose were “explanation” and “interaction with other road users”. Within the theme of explanation, the code “prior explanation” was mentioned 8 times and “a user manual” was mentioned 4 times. One participant mentioned they missed interactions with other road users.

These results provide some insight into what aspects of the HMI contribute to how it was perceived and experienced by participants. Many of the elements mentioned were indeed manipulated in the design of the stimuli (see Section 2.2.2.1).

4 Discussion and conclusions

4.1 Results and Discussion

The results from the factor analysis were in good agreement with the previous study [1]. As before, we found high internal consistency and identified two factors, namely a Perception and a Comprehension dimension. This demonstrates the robustness of the method, as it yields highly similar results when moving from a simple desktop setup to a more realistic, immersive driving simulator environment. The questionnaire was used with no time pressure, non-critical scenarios and ample opportunity to observe the HMI.

In the design of the stimuli, eight scenario versions were developed, each with three quality levels that were named bad, average, and good. This was done in several iterations involving multiple TNO project team members, reaching consensus over the designs that were used in the study. However, this was not validated independently. Thus, if, in a few scenarios, results do not show the expected effect of HMI quality, it is unclear whether this is due to the quality of the stimuli or a deficiency in the methodology. Still, with the majority of the results being in line with the expectations (i.e., improving with an improvement in the (designed) quality of the HMI), the first explanation seems more likely. For future work in this area, it is recommended to include an independent validation or the quality levels of the designs.

Many HMI design parameters were varied to obtain these quality levels, though not on perception-related parameters like font size, symbol size or contrast. This was reflected in the results: the Comprehension score showed significant variation, whereas most Perception scores were above +2, suggesting that most HMIs were easy to perceive, regardless of the HMI quality level. Nevertheless, the statistical analysis did reveal a significant main effect of HMI quality on the Perception score, primarily due to the bad HMI scoring lower than both the average and good HMIs. So, even when the scores are skewed towards the upper limit of the scale, the method is still sensitive enough to detect differences.

The distribution of the Comprehension score showed more variation, covering the entire range from -3 to +3, with a mode around +2. Here, the statistical analysis showed not only a significant main effect of HMI quality level, but also significant differences between bad-good, bad-average and average-good, all indicating that the Comprehension score increased as the HMI quality level improved. This pattern typically remained intact when zooming in on the individual scenario versions; only two out of eight versions deviated (showing either no effect or a reversed effect).

No main effect of experience with Level 2 systems was found on the Comprehension or Perception scores. The results did show an interaction effect, suggesting that some effects of HMI quality on the Perception score were found in the L2 experienced group but not in the L2 inexperienced group.

In the responsibility questions, it was seen that as the HMI quality level increased from bad to average to good, the percentage of correct answers either increased or did not change significantly. Therefore, better HMIs are interpreted correctly more often than poorer HMIs.

This trend was also observed in the regression models, which analysed the percentage of correct answers as a function of the Comprehension and Perception scores. In two out of eight regression models, the regression coefficients differed from zero, showing that higher Comprehension or Perception scores were associated with higher percentages of correct answers. However, in most regression models (six out of eight), the coefficients were not significantly different from zero, suggesting no effect of Comprehension or Perception scores on the percentage of correct answers. This may be due to a ceiling effect, as was the case for the questions on awareness and keeping hands on the steering wheel.

Identifying the elements that make an HMI clear and unambiguous was not within the scope of the project. Still, the results from the exit interview provide some important insights into this aspect. Many of the themes and codes matched the HMI elements that were manipulated in the design of the stimuli (see Section 2.2.2.1). Across the different participants, some central themes arose. With regards to clarity, participants emphasized the importance of supporting text and a good choice of colour schemes to present clear and concise information. Simplicity and auditory cues also played a noticeable role in how the users perceived the dashboards. Uncertainty and unclarity often arose from an overload of information. There appears to be a clear connection between the factors that contributed to clarity and certainty. For example, the colour green was associated with clarity as well as with certainty. Similarly, there was a connection between the factors that led to unclarity and uncertainty.

4.2 Conclusions and Recommendations

In this project, the questionnaire was tested on both transitioning and static HMIs, covering various levels of automation. These ranged from manual driving, through L2 ADAS, to partial driving automation (L3), where the driver can disengage but must be ready to take over. Overall, this methodology was found to be capable of detecting differences among HMI quality levels across a variety of ADAS and ADS scenarios. This strengthens the conclusion from the previous project [1, 2] that this questionnaire can measure whether the information from an ADS HMI is clear and unambiguous to its user.

No overall effects of experience with L2 systems for the participants were observed. It was only found that for the Perception score, the participants with L2 experience were more sensitive to the quality differences in the HMI than those without L2 experience. Future research could validate this finding by again having groups of participants with and without L2 experience. Alternatively, if this factor cannot be varied, the safest approach would be to have only participants with L2 experience.

Limitations of this approach and suggestions for further development remain in line with previous reports [1, 2]. That is: this questionnaire showing “clear and unambiguous” information cannot be considered as sufficient to guarantee safe responses by the user. Even with “clear and unambiguous” information, a user may respond incorrectly or too late. Assessing whether responses are timely and correct requires the method to be extended, by introducing dynamic scenarios for participants in (either real or simulated) traffic situations.

Concerning further development of the Clear and Unambiguous Questionnaire, we recommend the following steps (evolving the recommendations of [2]).

- Development of a set of benchmark HMIs, which can function as a reference points for questionnaire results on the HMI under evaluation. Such a set should preferably be developed by a larger community, e.g., involving HMI/UX/HF experts from academia, research organisations, and OEMs. This effort could focus on developing various HMIs with a clear consensus on their quality level, including good as well as (unacceptably) bad HMIs, for various levels of automation. These can then serve as references for assessing new HMIs and for defining thresholds.

- Development of a test procedure that includes measuring user responses to HMI information. This should cover various traffic scenarios, including potentially critical ones. Measurement of user responses should involve all aspects that determine safety, including response timing and type of response (e.g., braking vs. steering). In addition, the time required by the user to perceive and interpret HMI information should be measured as part of the response time when possible.
- The test scenarios should also cover situational or environmental factors that may negatively affect perception and comprehension of HMI information. These include for instance lighting conditions, acoustic noise, non-driving related activities and passenger activity. Since combining these factors quickly leads to a very large number of test cases, first a prioritization of test cases defined by these factors needs to be established.
- Finally, the most important dimensions of user diversity need to be identified. The current study explored include familiarity with ADAS, but factors like ranges of perceptual and cognitive abilities in the (potential) user population still need to be addressed.

5 References

[1] J. Souman, R. Post, J. Hogema and M. Hoedemaeker, “Development of a measurement scale to assess Automated Driving System HMIs (TNO rapport 2023 R11506),” Helmond: TNO Molbity and Built Environment, 2023.

[2] J. Hogema, R. Post, R. Verstegen and J. Souman, “Automated Driving System HMIs: "Clear and Unambiguous"?,” *16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI ’24),* p. 10, 2024.

[3] M. Perrier, T. Louw, Carsten and O., “User-centred design evaluation of symbols for adaptive cruise control,” *Cognition, Technology & Work,* vol. 23, pp. 685-703, 2021.

[4] SAE, “J3016. The Principles of Operation Framework: A Comprehensive Classification Concept for Automated Driving Functions.,” 2021.

[5] M. de Goede, R. Jansen and E. van Grondelle, “User-centred design for type approval of AD(A)S. Roadmap towards a process audit (R-2022-16),” Stichting Wetenschappelijk Onderzoek Verkeersveiligheid (SWOV), The Hague, Netherlands, 2023.

[6] V. Braun and V. Clarke, “Using Thematic Analysis in Psychology,” *Qualitative Research in Psychology,* vol. 3, no. 2, pp. 77-101, 2006.

Appendix A: HMI Designs and Scenario Simulations

The HMI implemented in the driving simulator consisted of three main elements: a visual display in the dashboard, audio, and LEDs in the rim of the steering wheel.

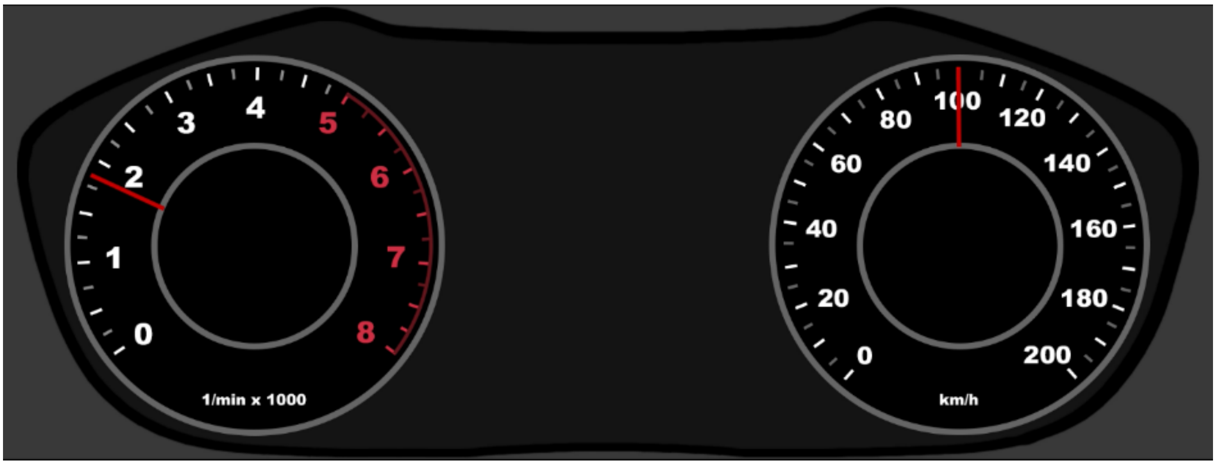
A.1 Dashboard

This section shows the dashboard part of the HMI. In static scenarios (1-12 and 25), only one image was shown on the HMI and hence only one picture is shown below. For the transition scenarios (13-24 and 26), several images are shown:

- The initial dashboard image;
- For transitions: the dashboard image, or if blinking elements were used: several images;
- The final dashboard image.

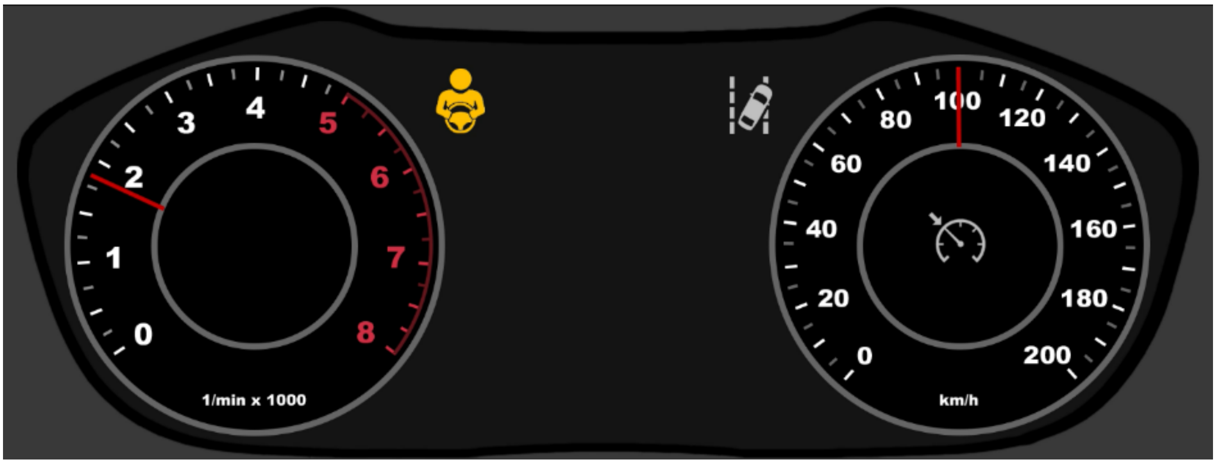
Scenario 1

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
1	Good	Static	A	L0	Cruise Control Off + Lane Departure Warning on OR active lane departure assist	Middle



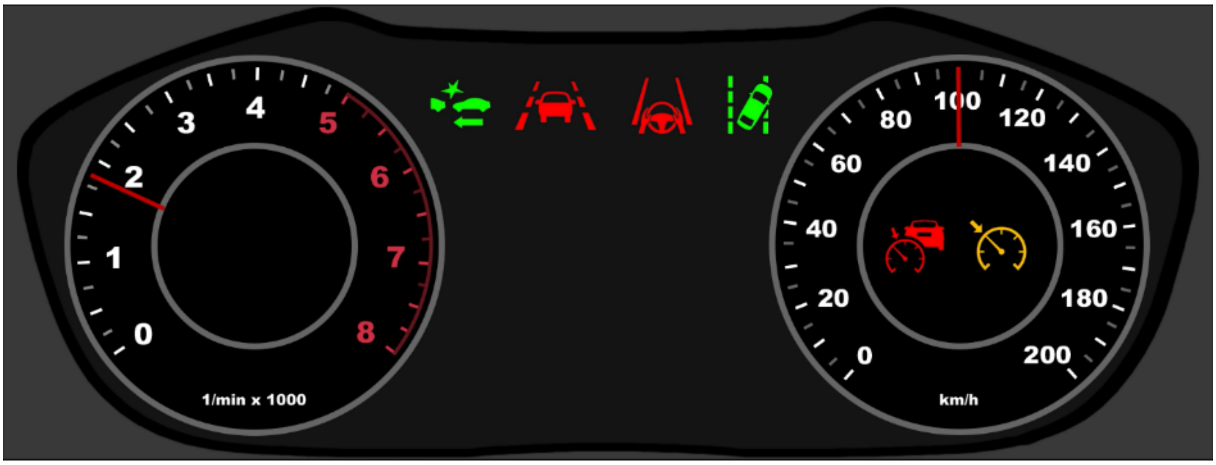
Scenario 2

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
2	Average	Static	A	L0	Cruise Control Off + Lane Departure Warning on OR active lane departure assist	Left



Scenario 3

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
3	Bad	Static	A	L0	Cruise Control Off + Lane Departure Warning on OR active lane departure assist	Right

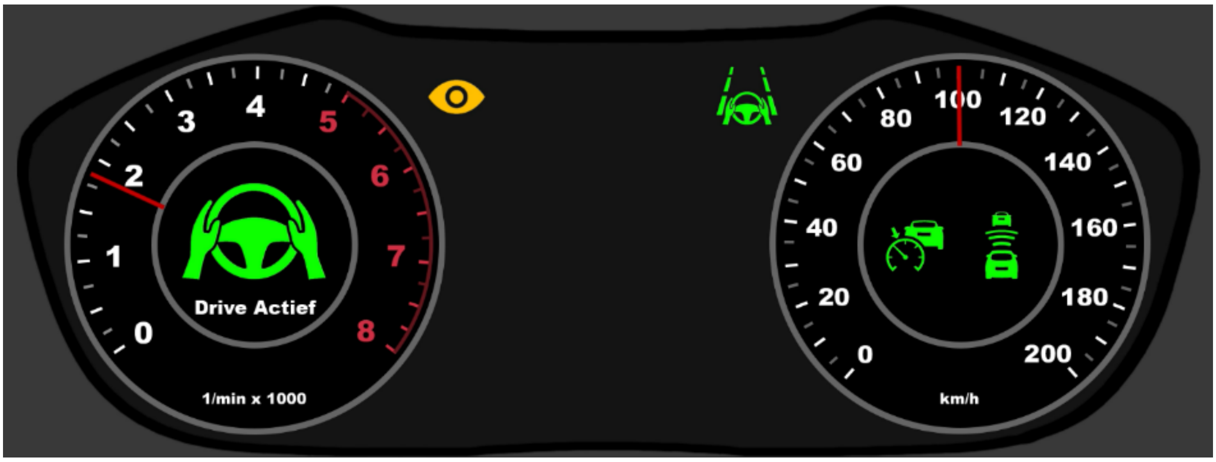


Scenario 4

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
4	Good	Static	B	L2 Hands-On	Active	Left

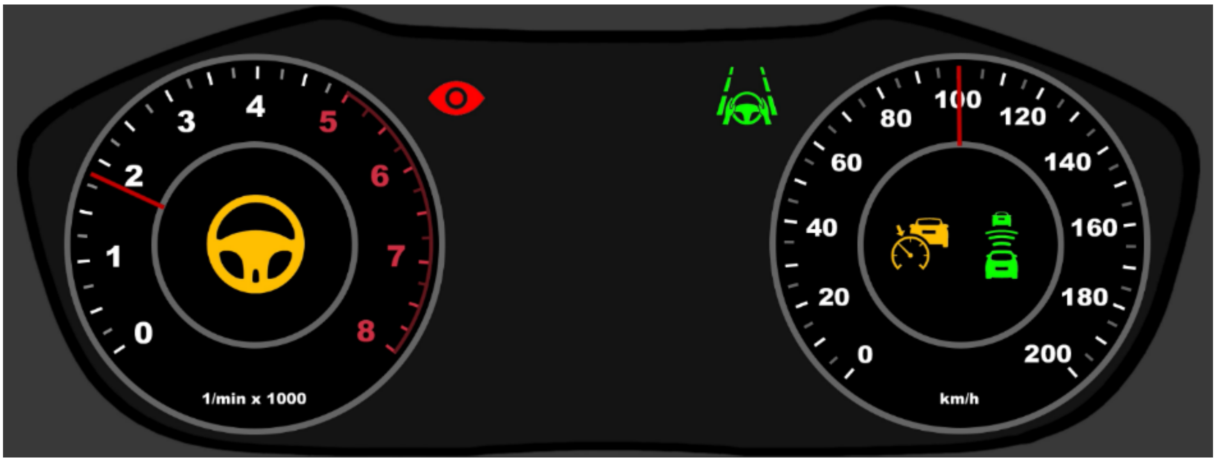
Scenario 5

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
5	Average	Static	B	L2 Hands-On	Active	Right



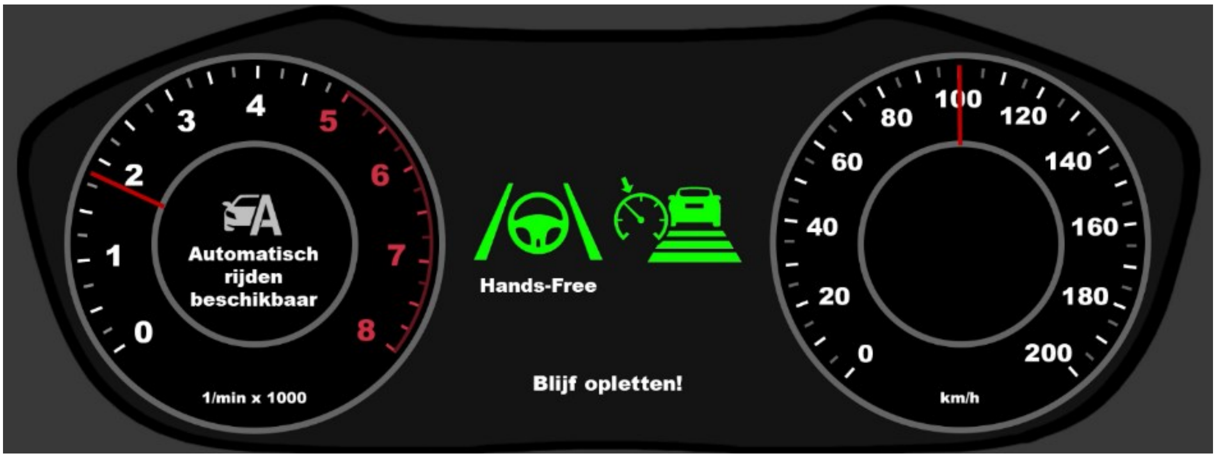
Scenario 6

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
6	Bad	Static	B	L2 Hands-On	Active	Middle



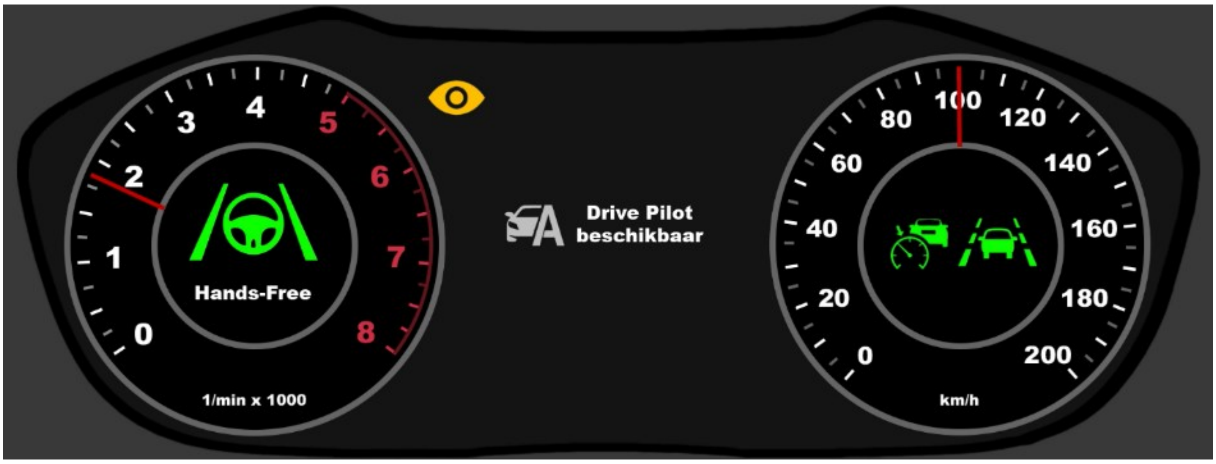
Scenario 7

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
7	Good	Static	C	L2-Hands-Off	Active + L3 Available	Right



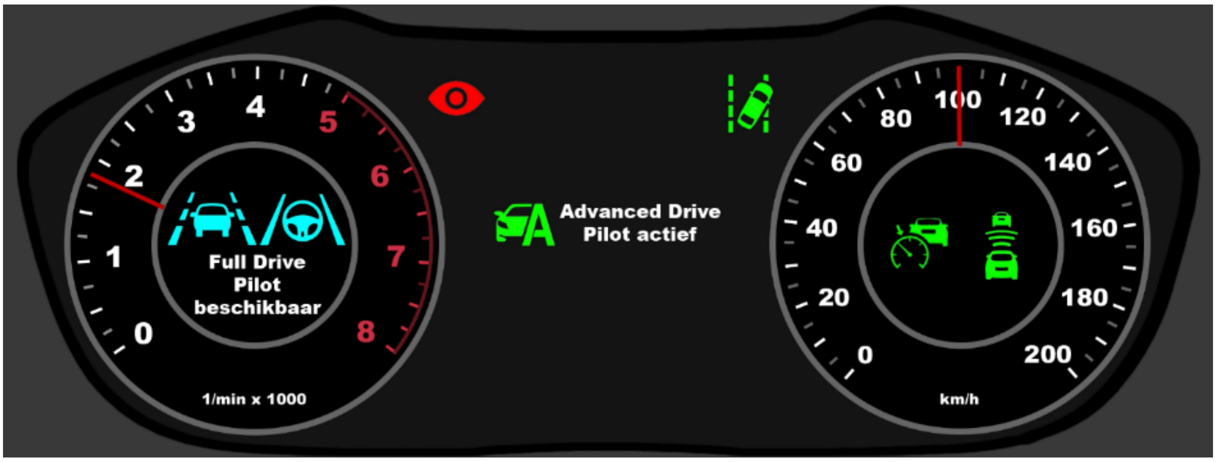
Scenario 8

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
8	Average	Static	C	L2-Hands-Off	Active + L3 Available	Middle



Scenario 9

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
9	Bad	Static	C	L2-Hands-Off	Active + L3 Available	Left

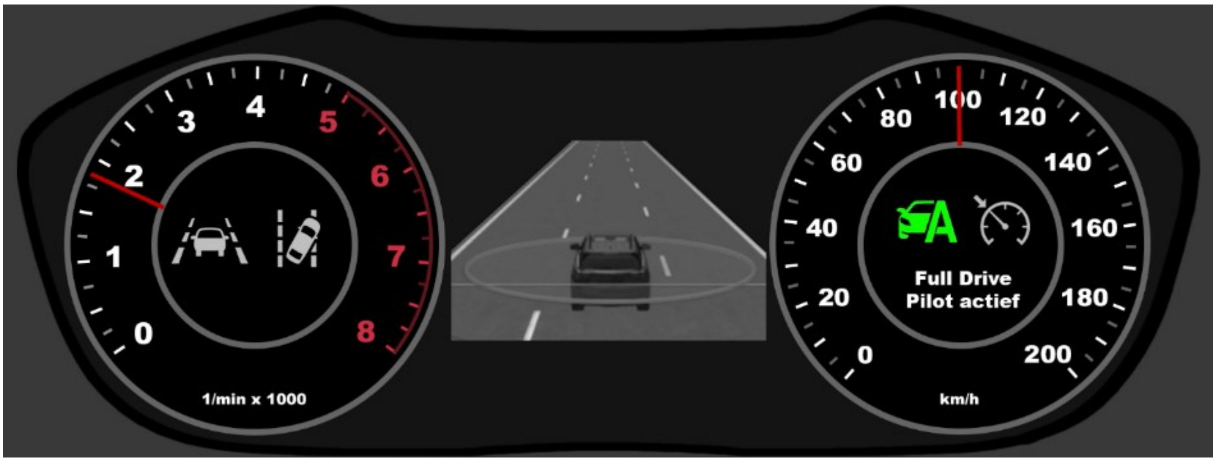


Scenario 10

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
10	Good	Static	D	L3	Active	Left

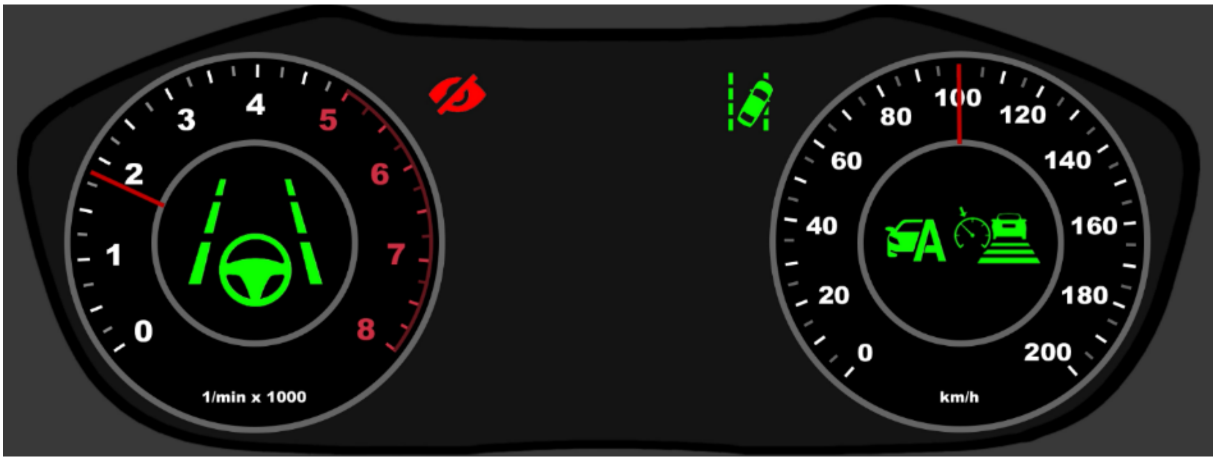
Scenario 11

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
11	Average	Static	D	L3	Active	Middle



Scenario 12

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
12	Bad	Static	D	L3	Active	Right

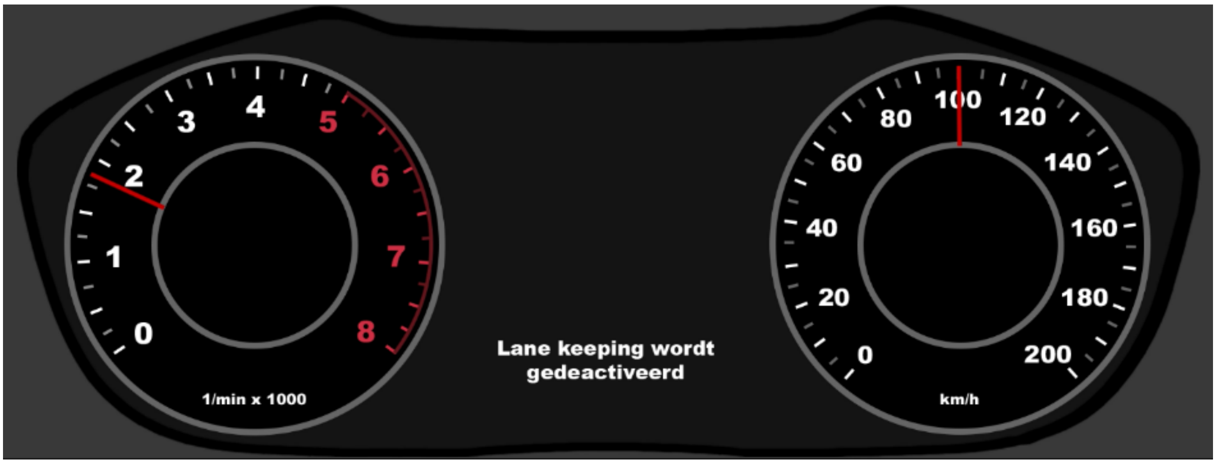


Scenario 13

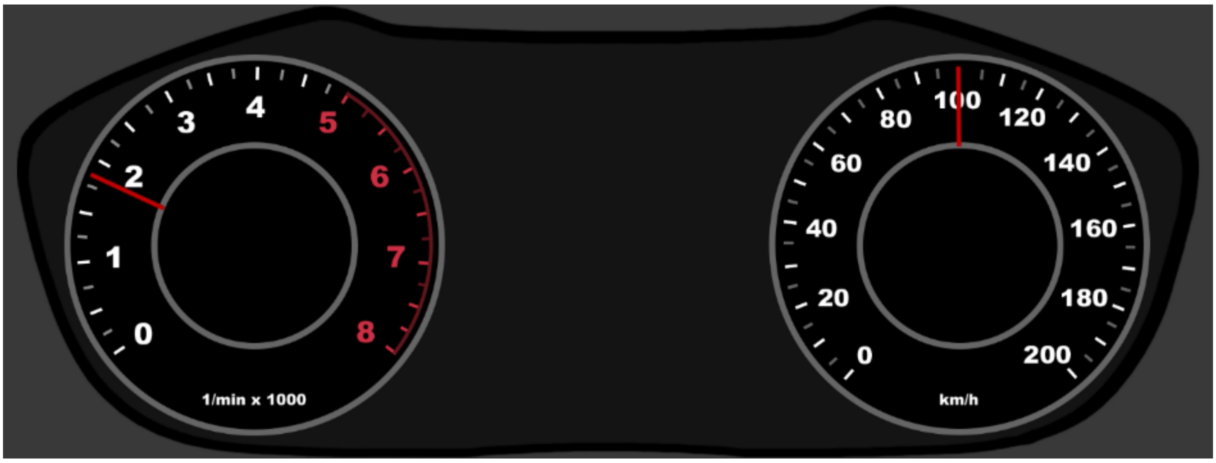
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
13	Good	Transition	E	L2 Hands-on to L0		Right

Start

Transition



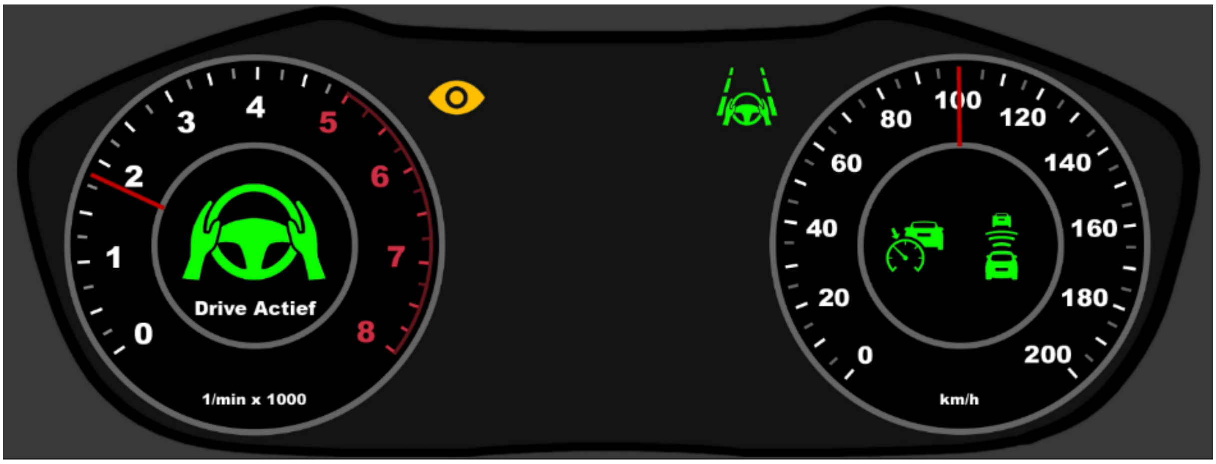
End



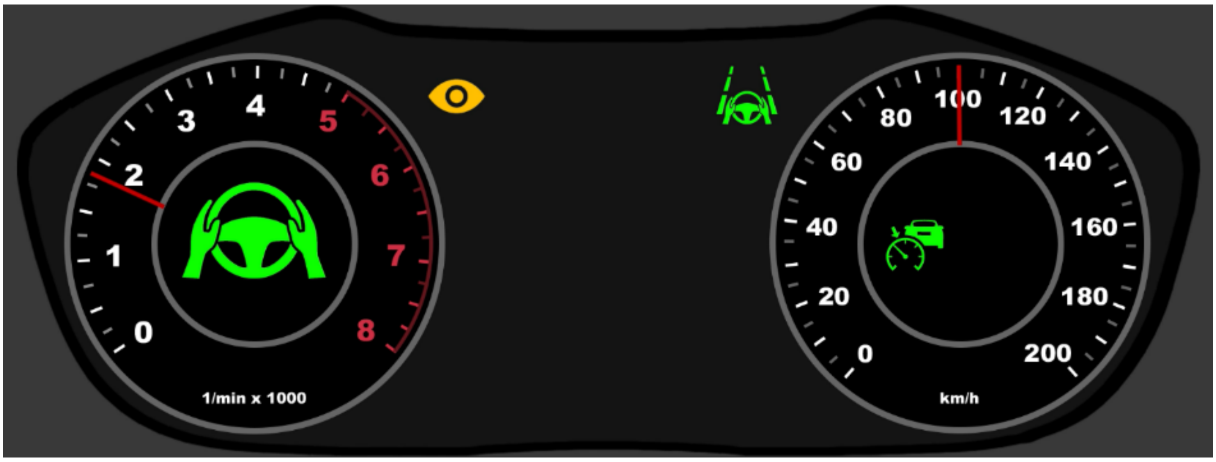
Scenario 14

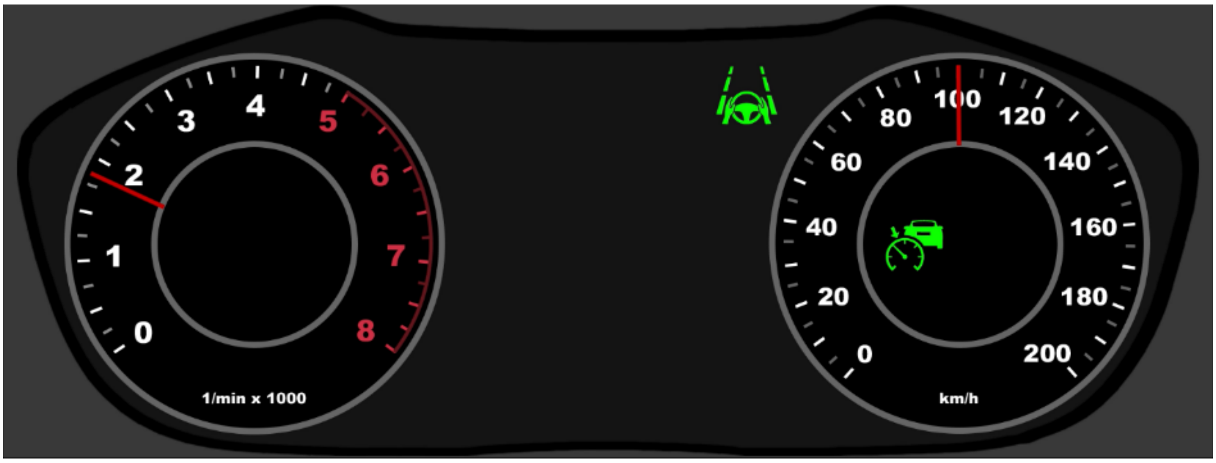
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
14	Average	Transition	E	L2 Hands-on to L0		Left

Start

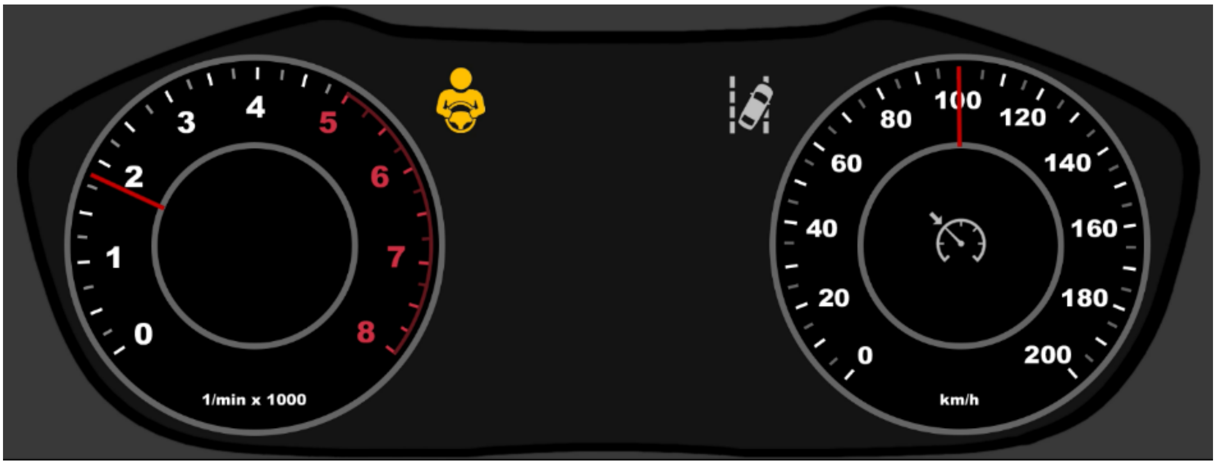


Transition





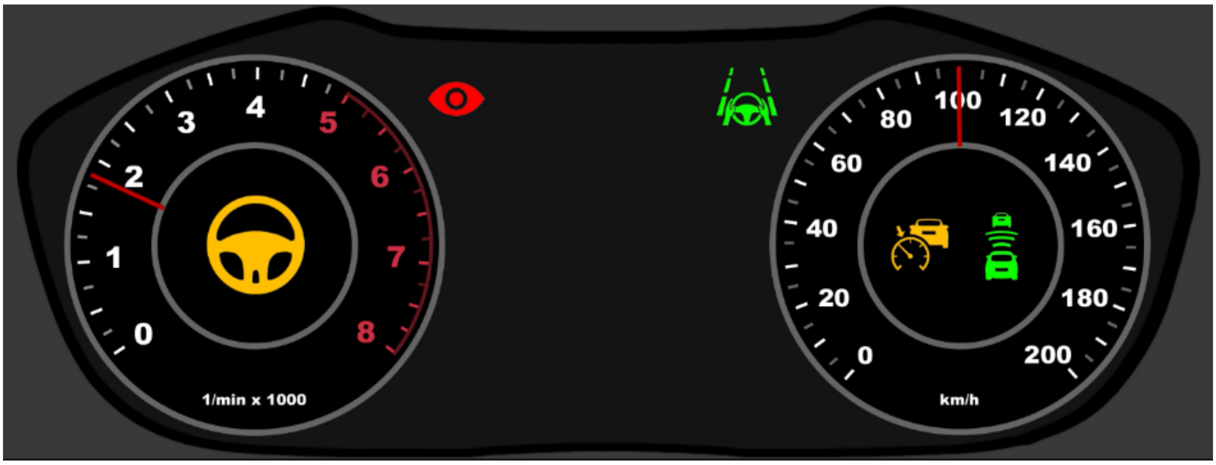
End



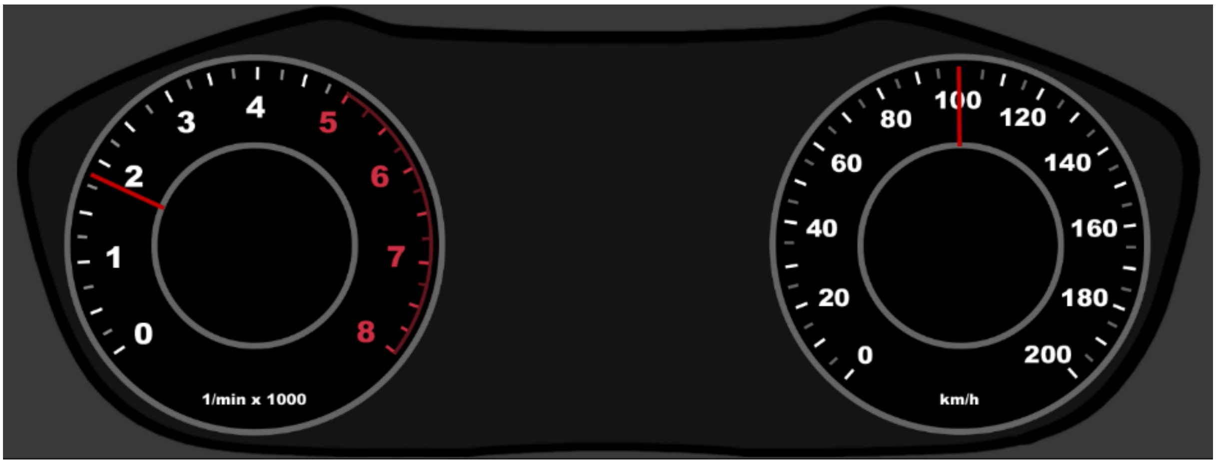
Scenario 15

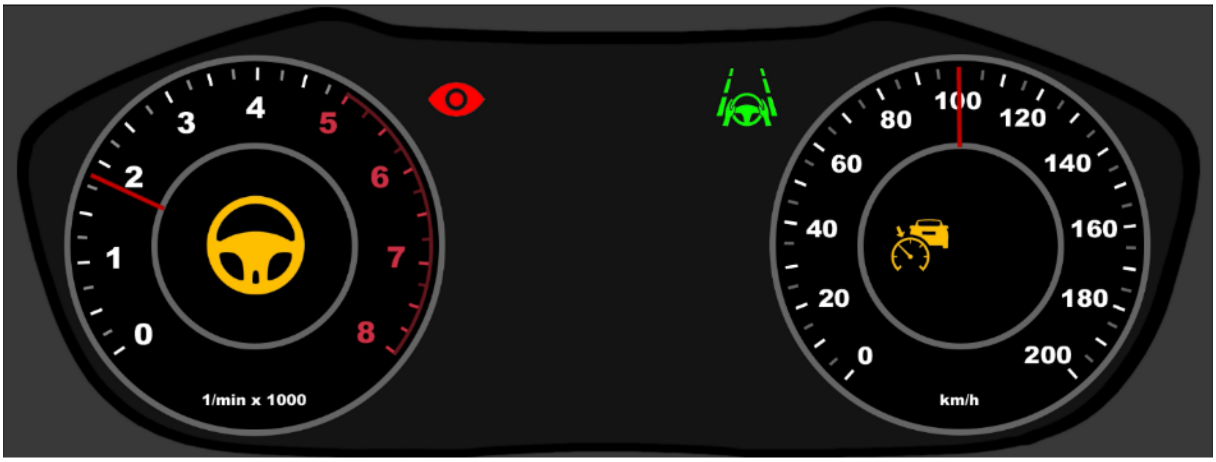
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
15	Bad	Transition	E	L2 Hands-on to L0		Middle

Start

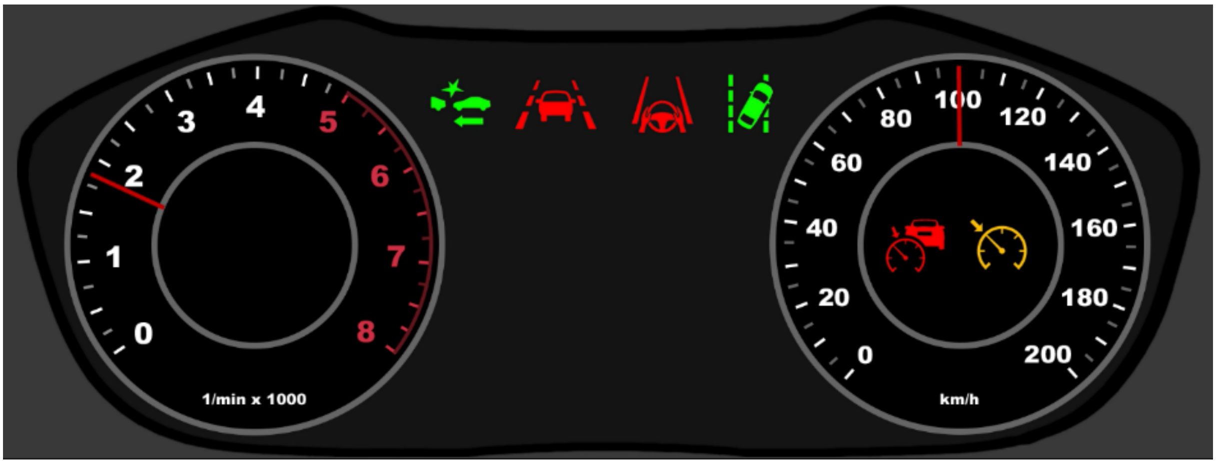


Transition





End

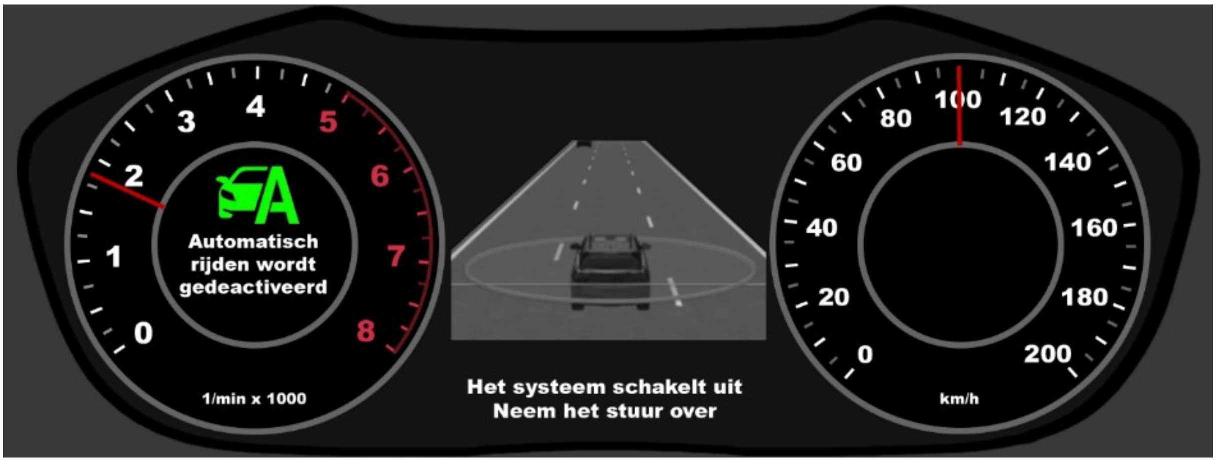


Scenario 16

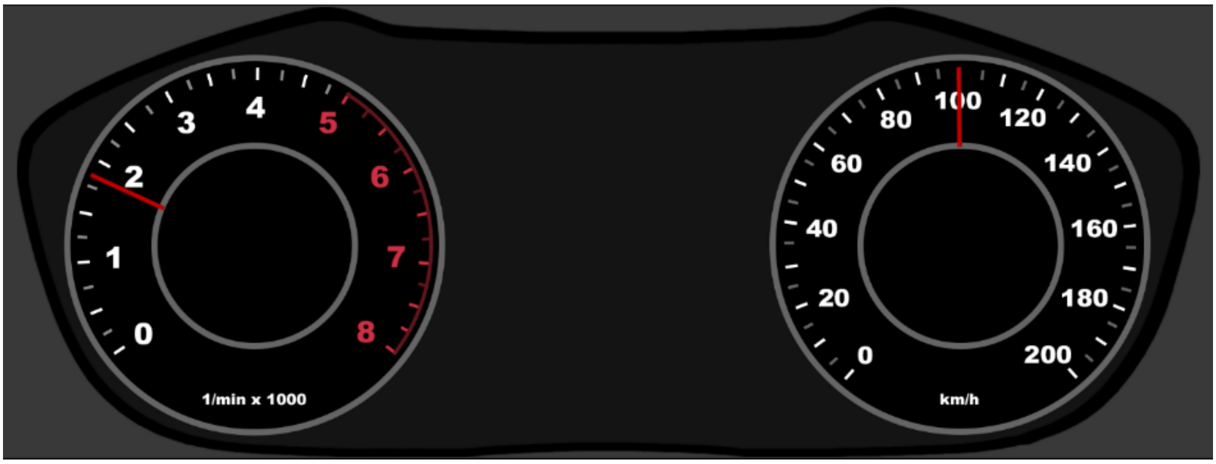
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
16	Good	Transition	F	L3 to L0		Middle

Start

Transition



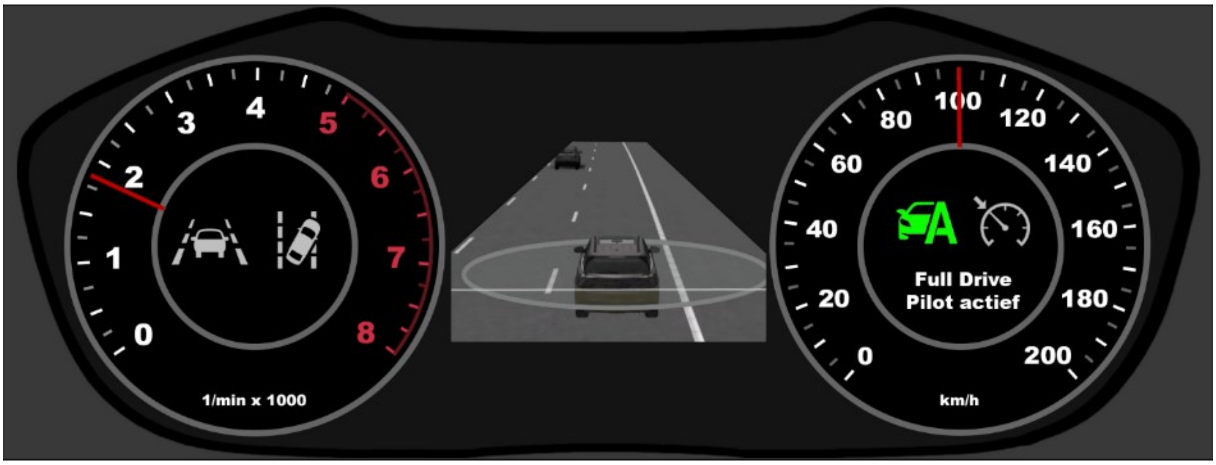
End



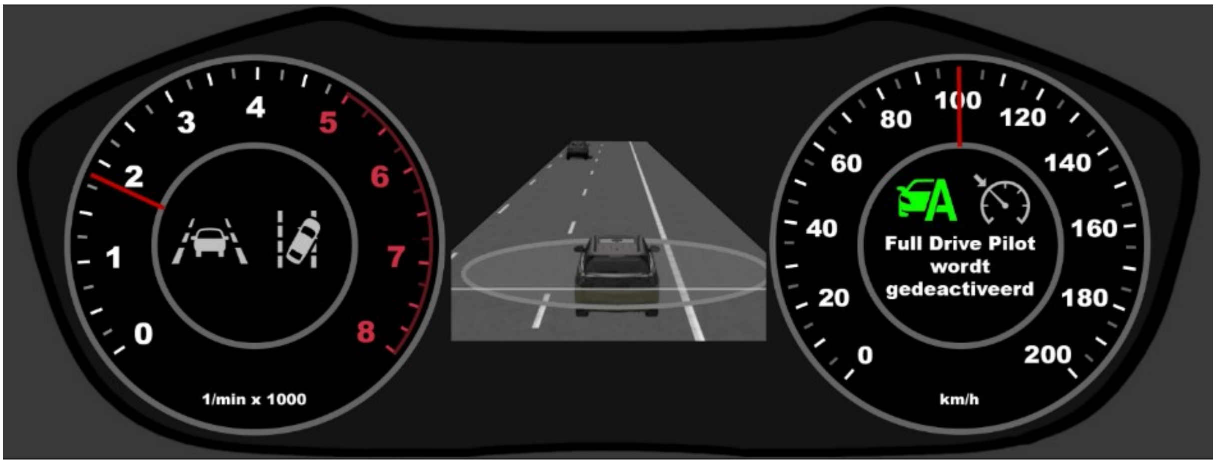
Scenario 17

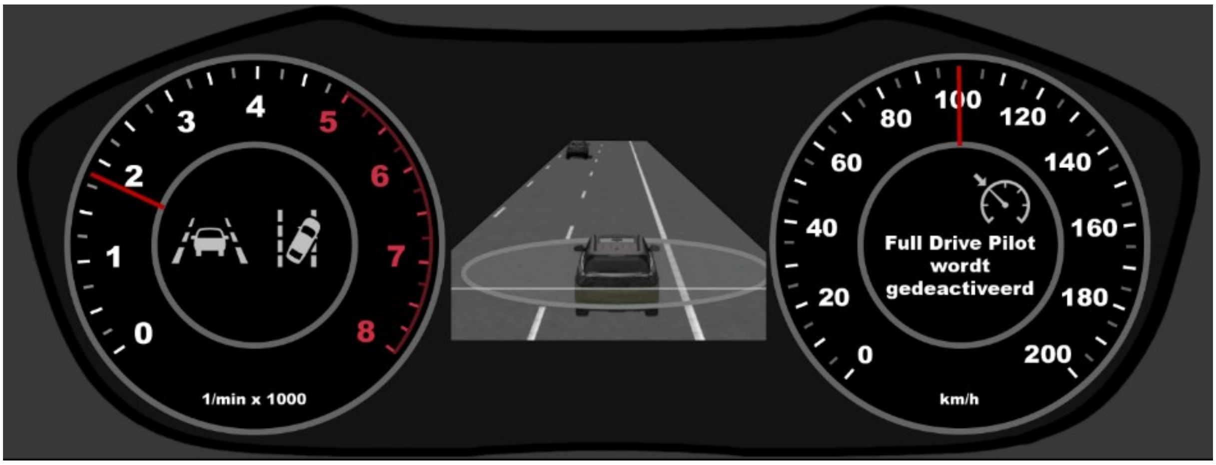
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
17	Average	Transition	F	L3 to L0		Right

Start

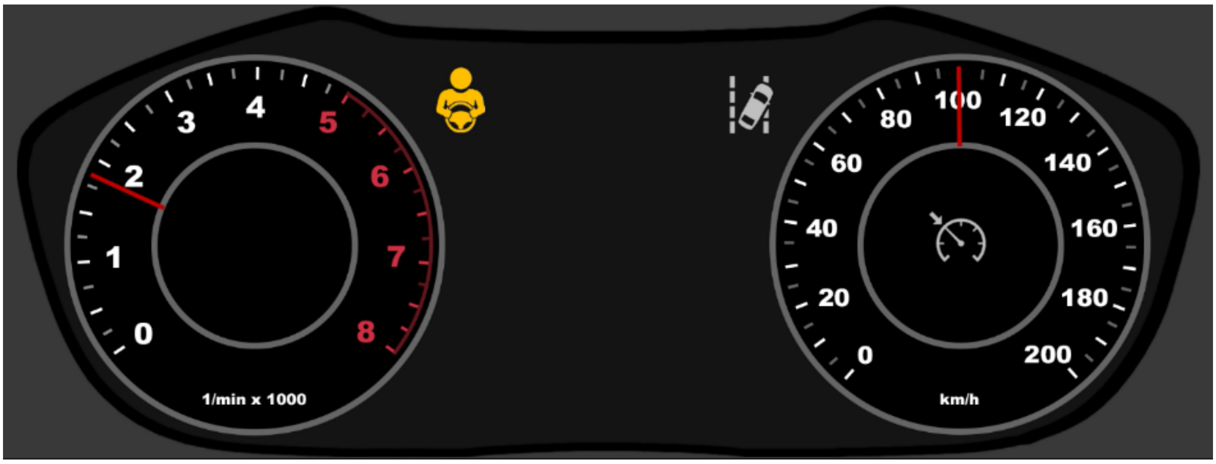


Transition





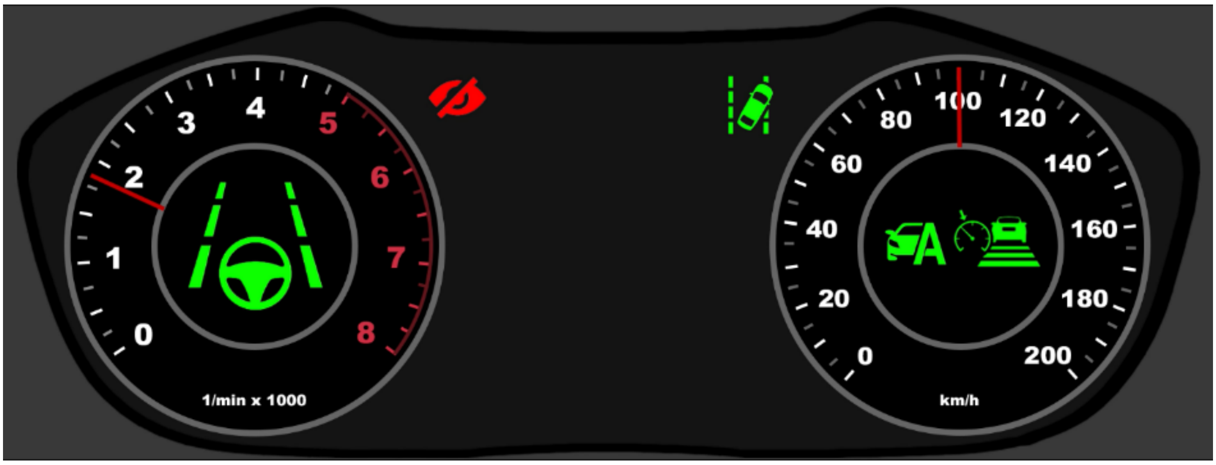
End



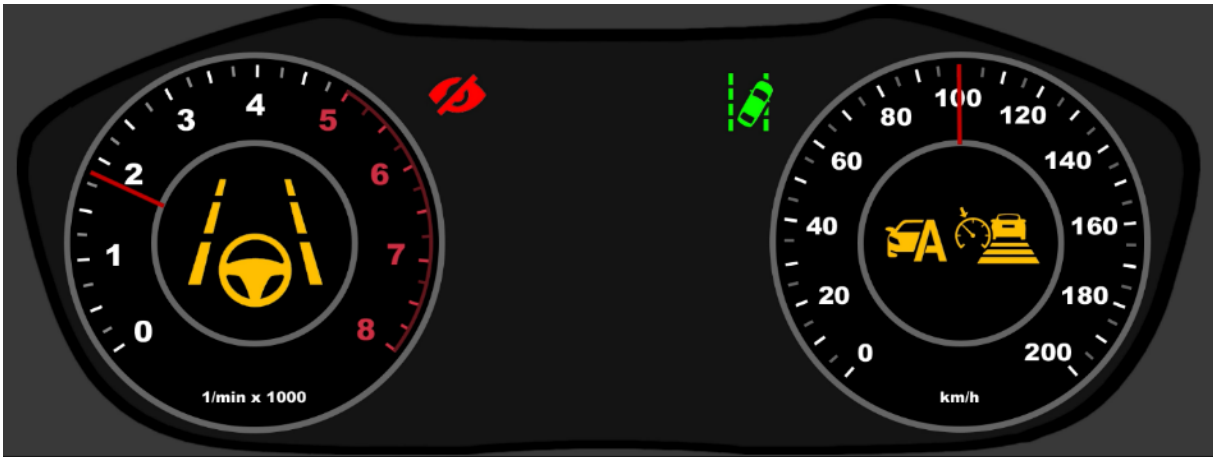
Scenario 18

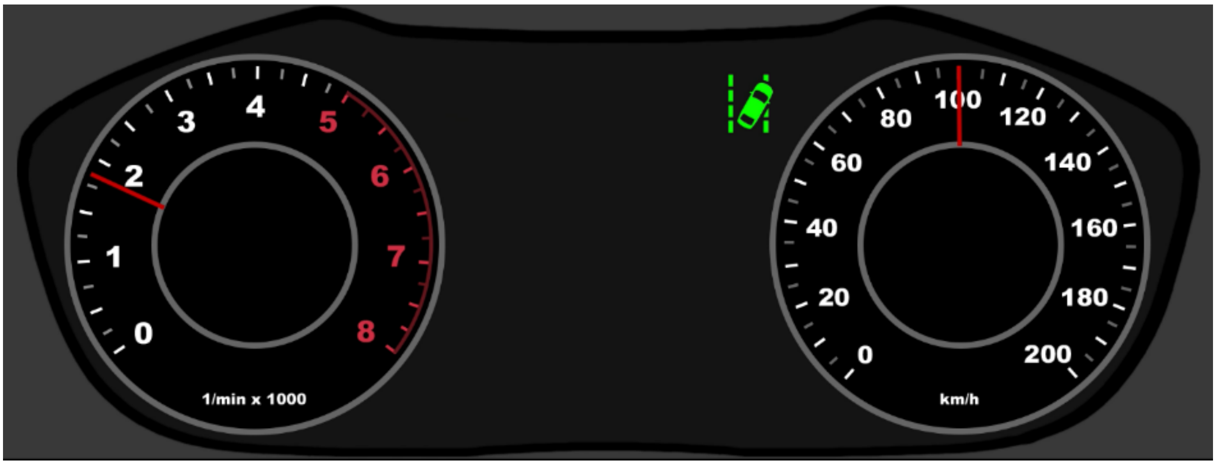
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
18	Bad	Transition	F	L3 to L0		Left

Start

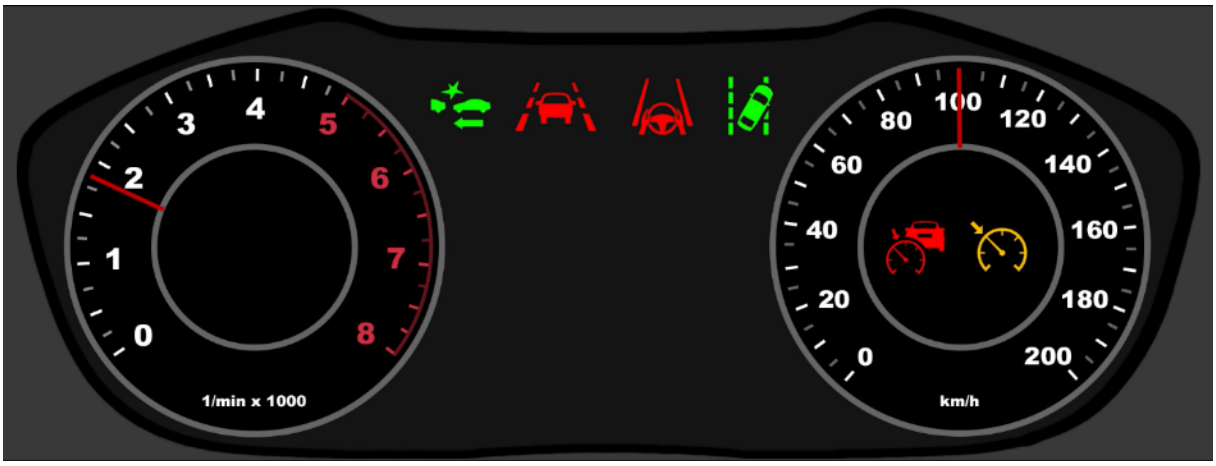


Transition





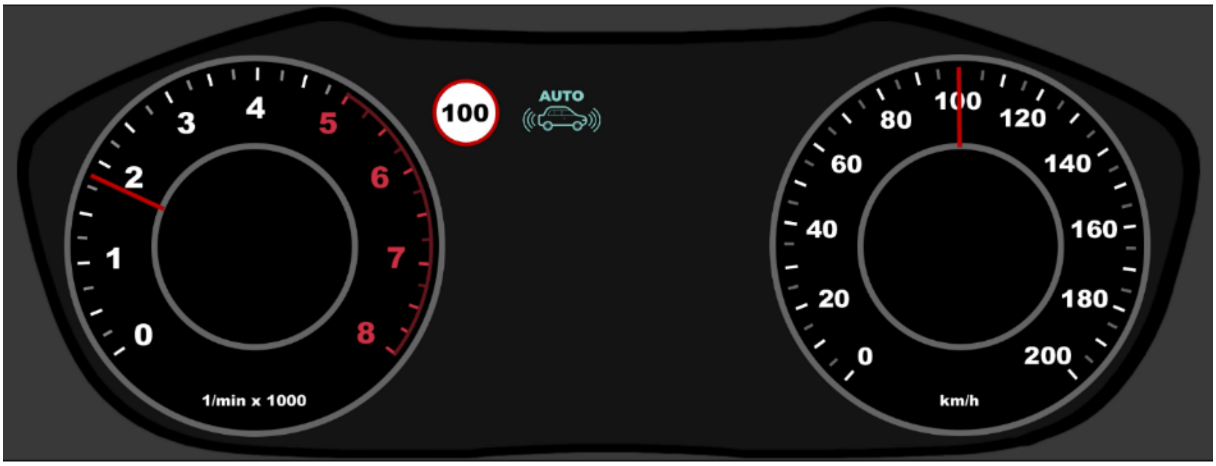
End



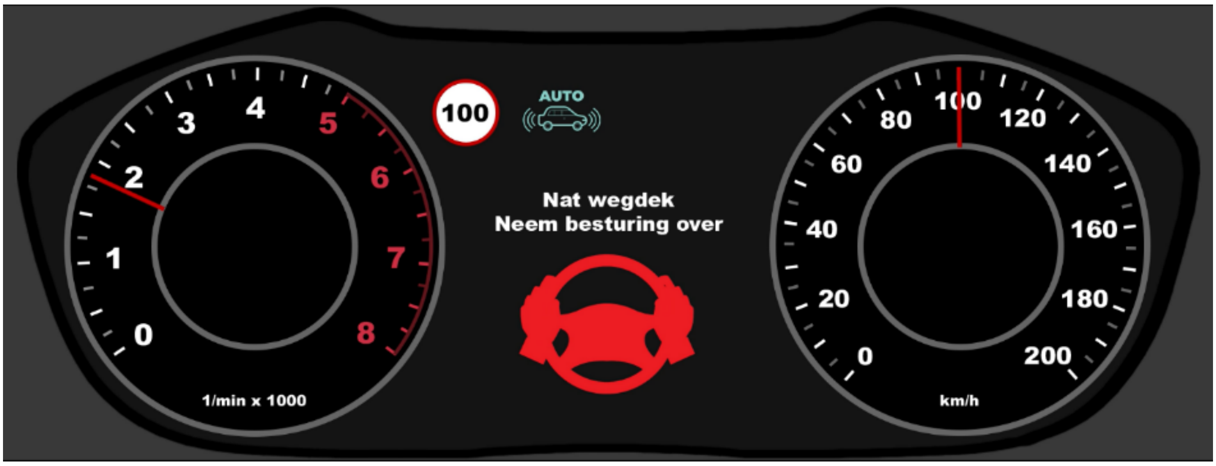
Scenario 19

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
19	Good	Transition	G	L3 to L0		Right

Start



Transition

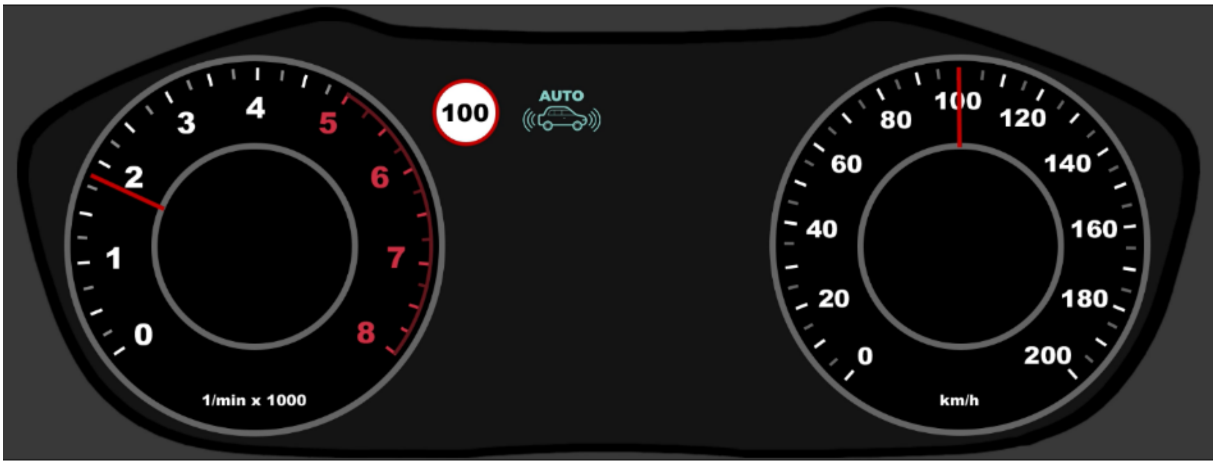


End

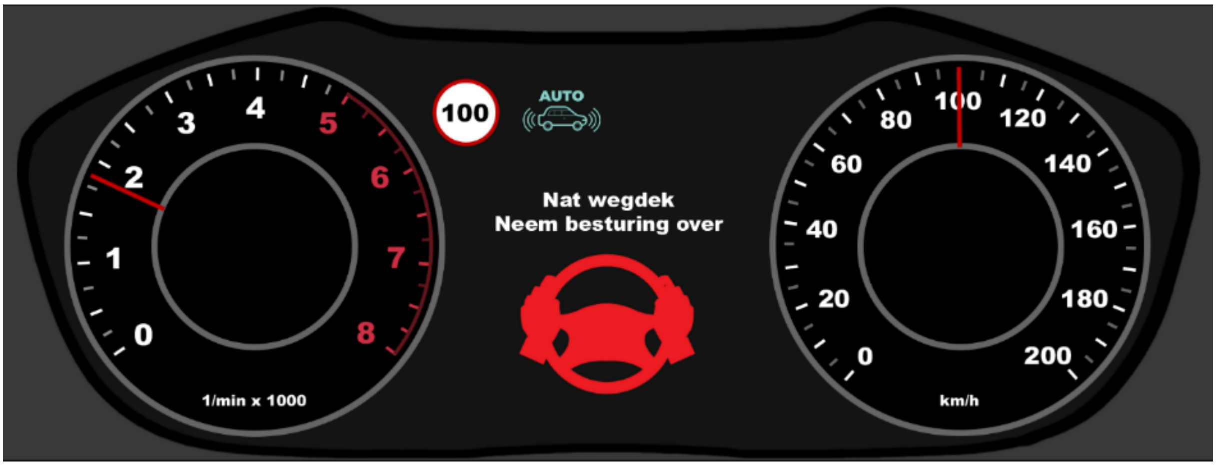
Scenario 20
Note: As explained in Section 2.2.2.1, Scenarios 19 and 20 were identical in terms of the HMI; only the lanes in the videos differed.

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
20	Average	Transition	G	L3 to L0		Left

Start



Transition

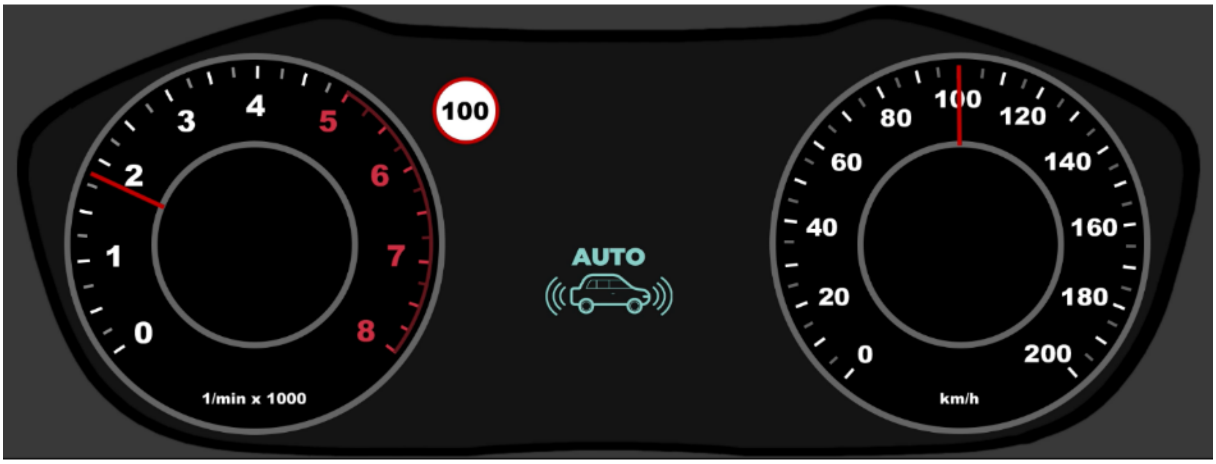


End

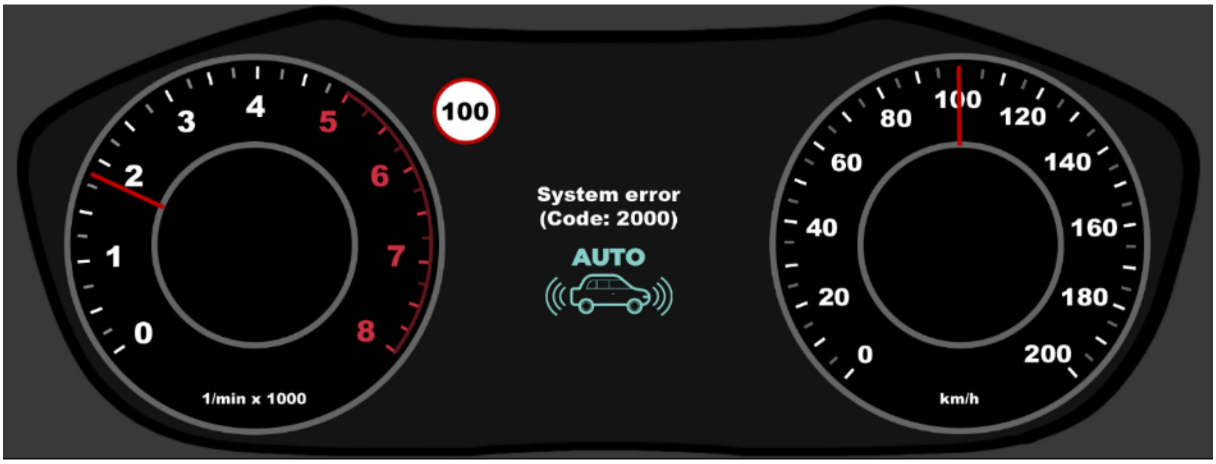
Scenario 21

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
21	Bad	Transition	G	L3 to L0		Middle

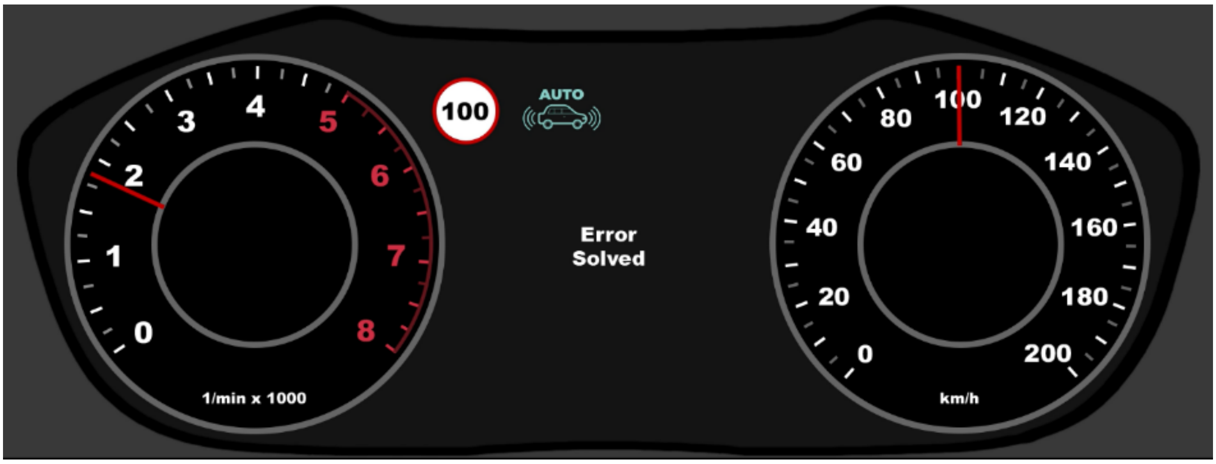
Start



Transition



End



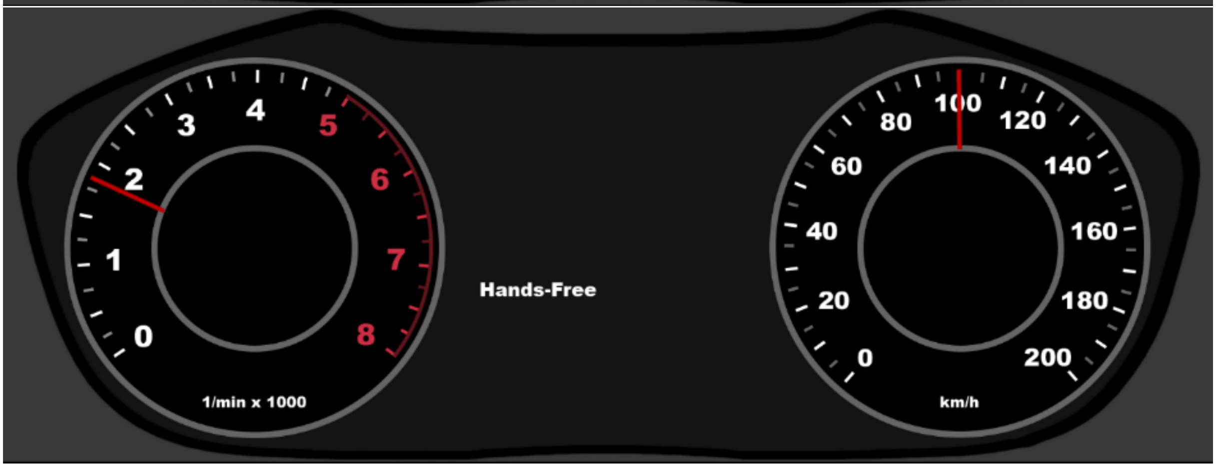
Scenario 22

Scenario ID	Quality	Type	Version	AD Level	Status	Lane
22	Good	Transition	H	L2 Hands-off to L0		Left

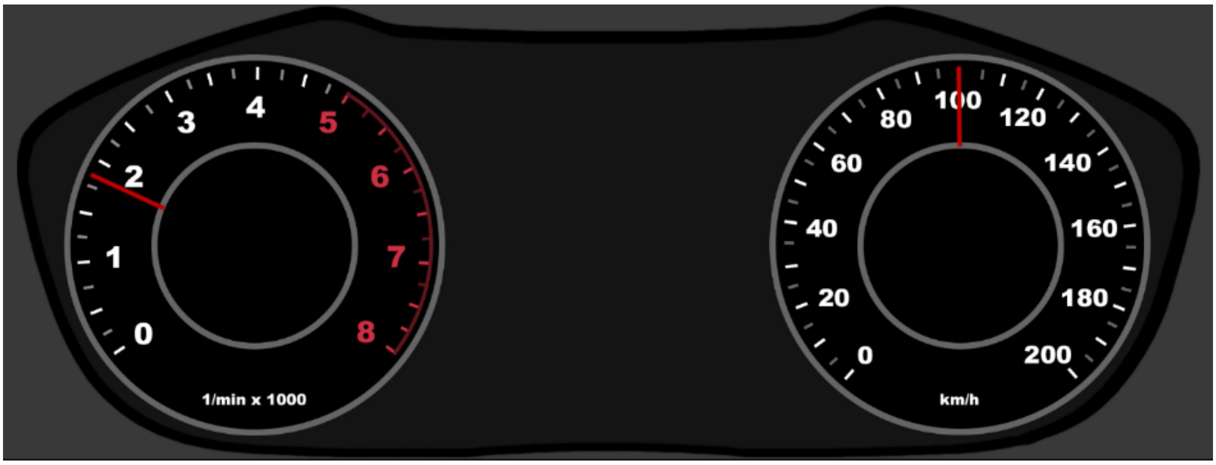
Start

Transition

Transition Escalation



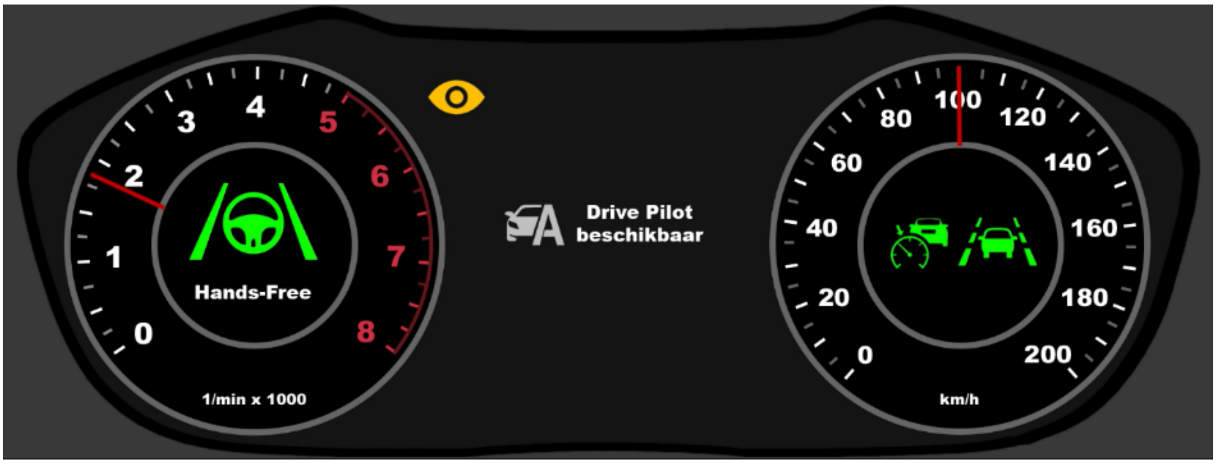
End



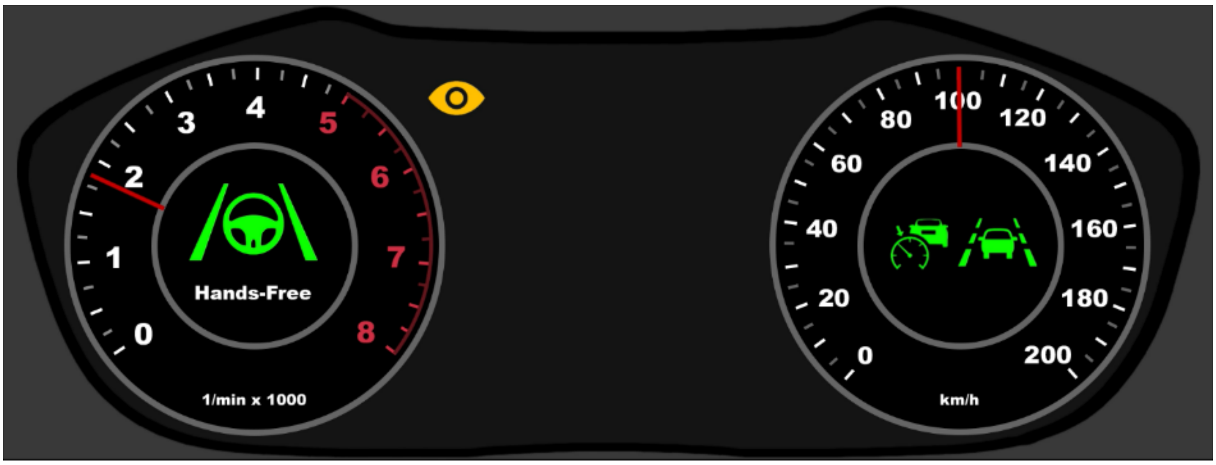
Scenario 23

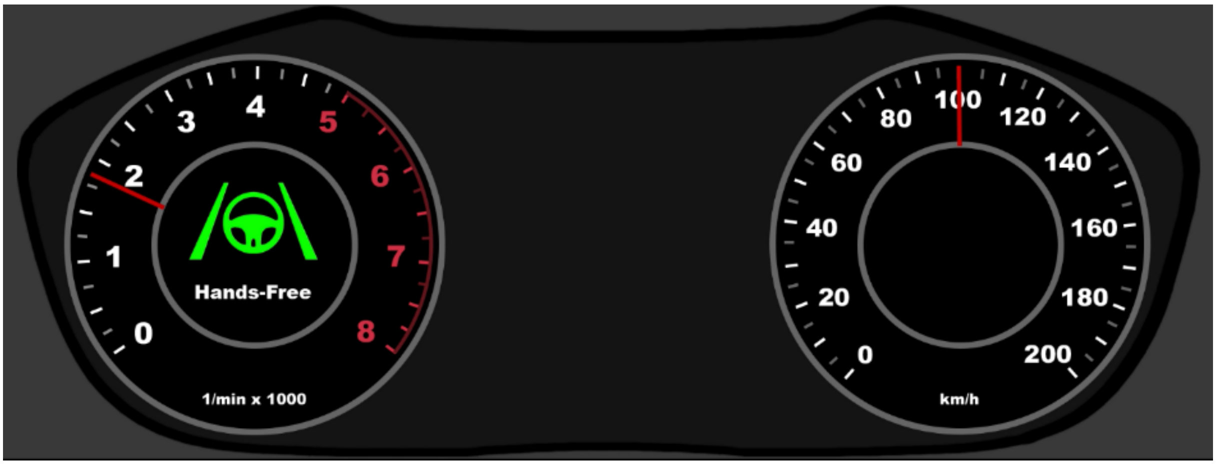
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
23	Average	Transition	H	L2 Hands-off to L0		Right

Start

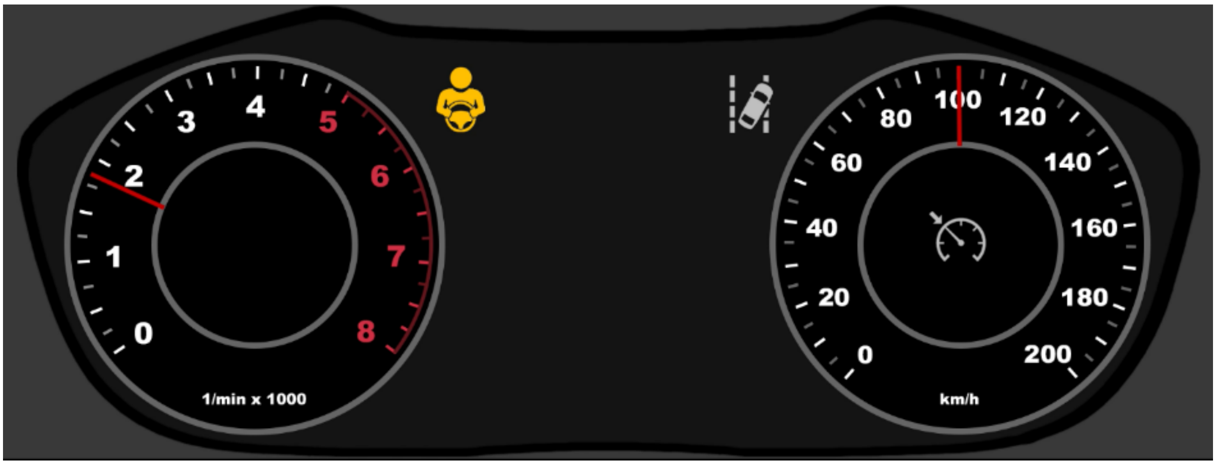


Transition





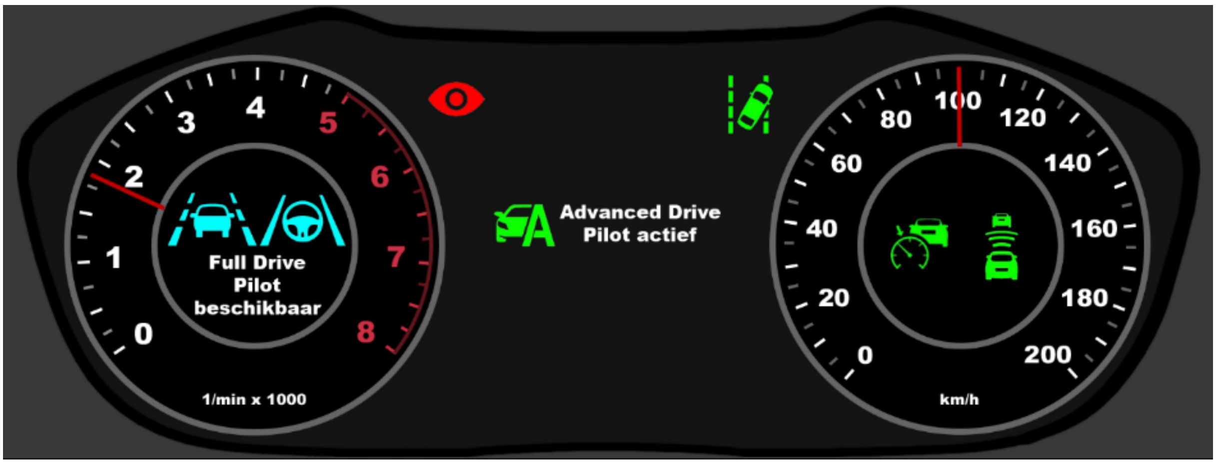
End



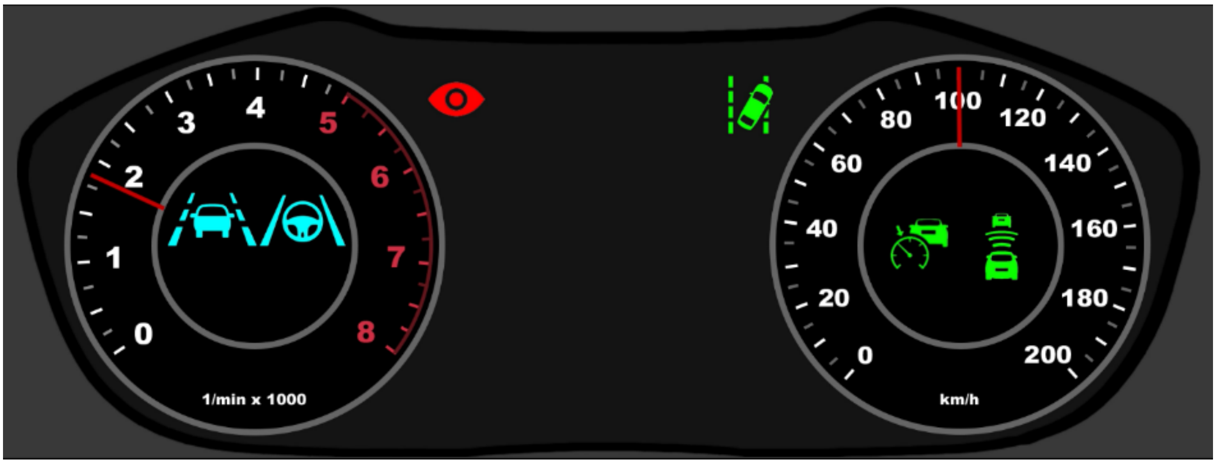
Scenario 24

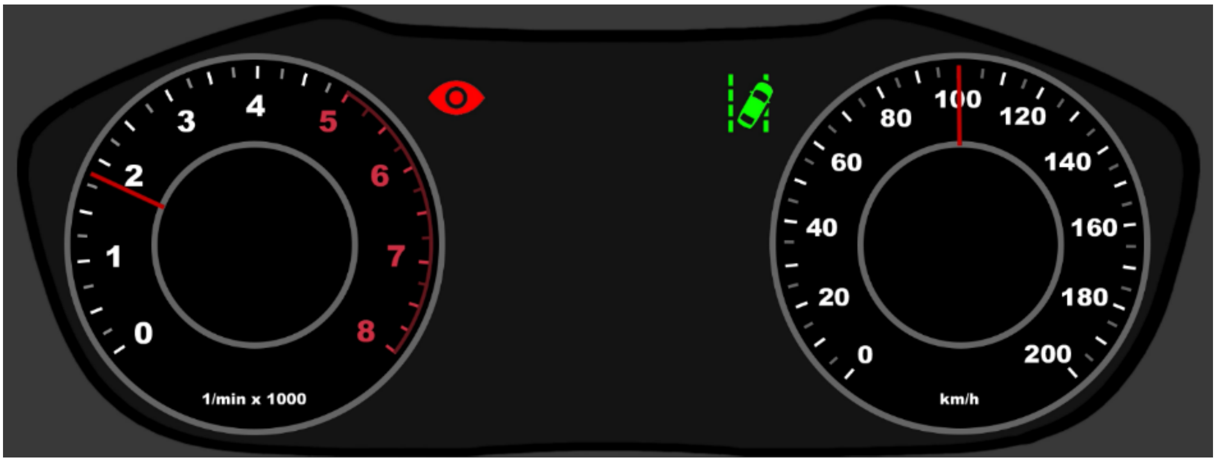
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
24	Bad	Transition	H	L2 Hands-off to L0		Middle

Start

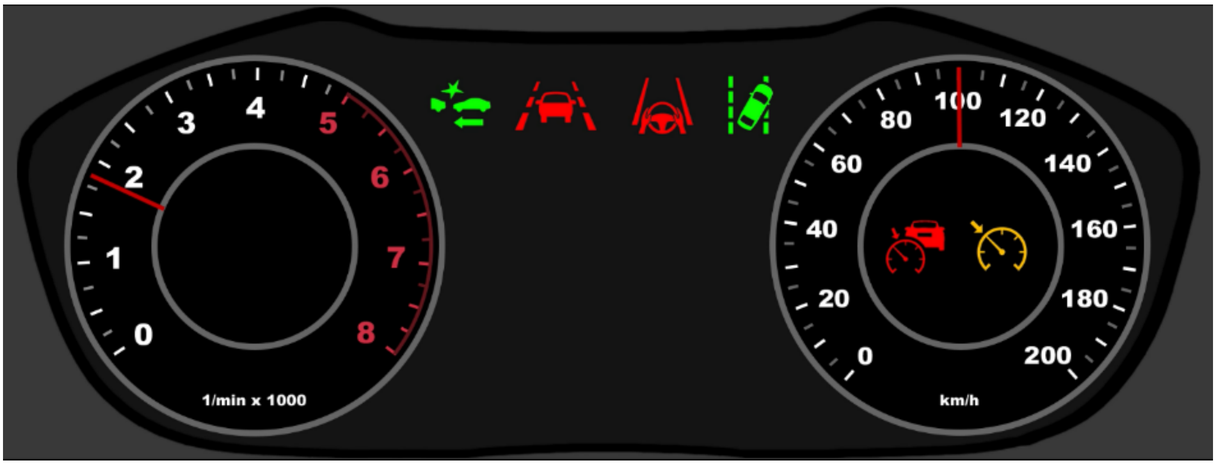


Transition



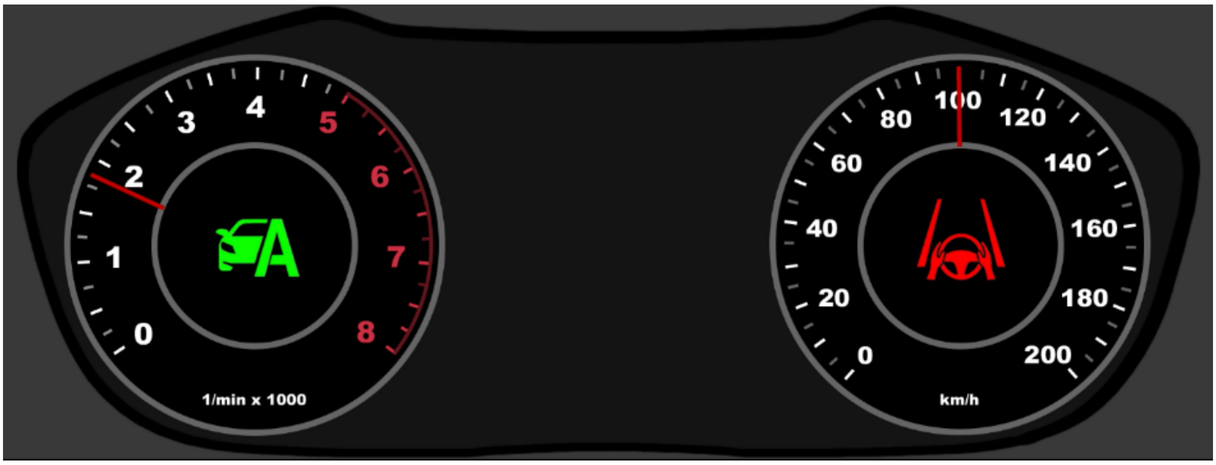


End



Scenario 25

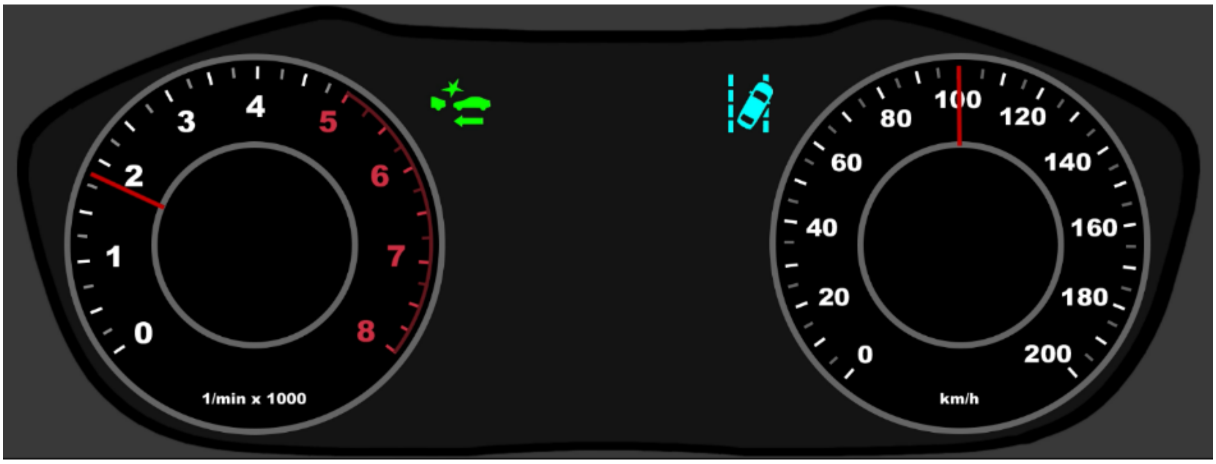
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
25	Average	Practice run, static		L3		Middle



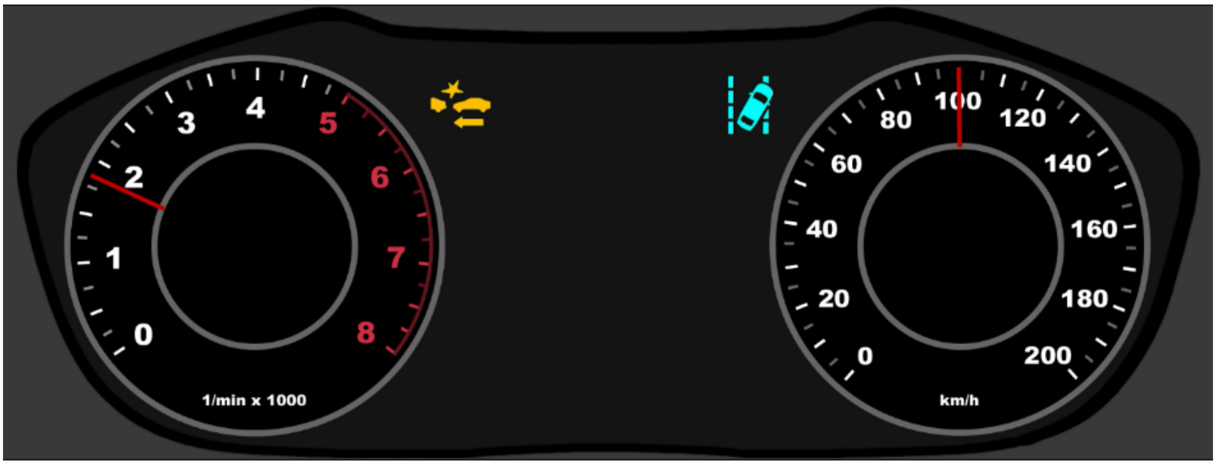
Scenario 26

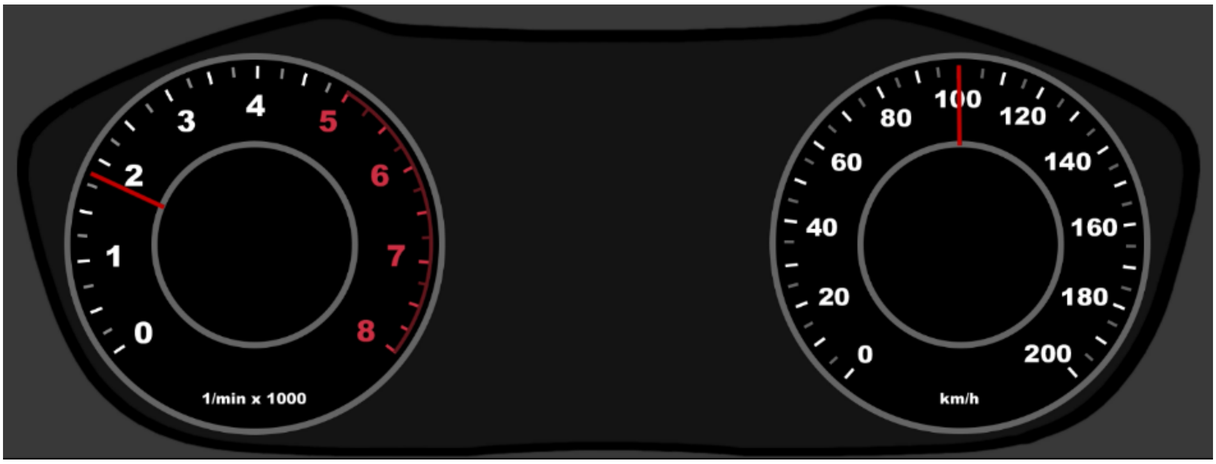
Scenario ID	Quality	Type	Version	AD Level	Status	Lane
26	Average	Practice run, Transition		L1 → L0		Left

Start

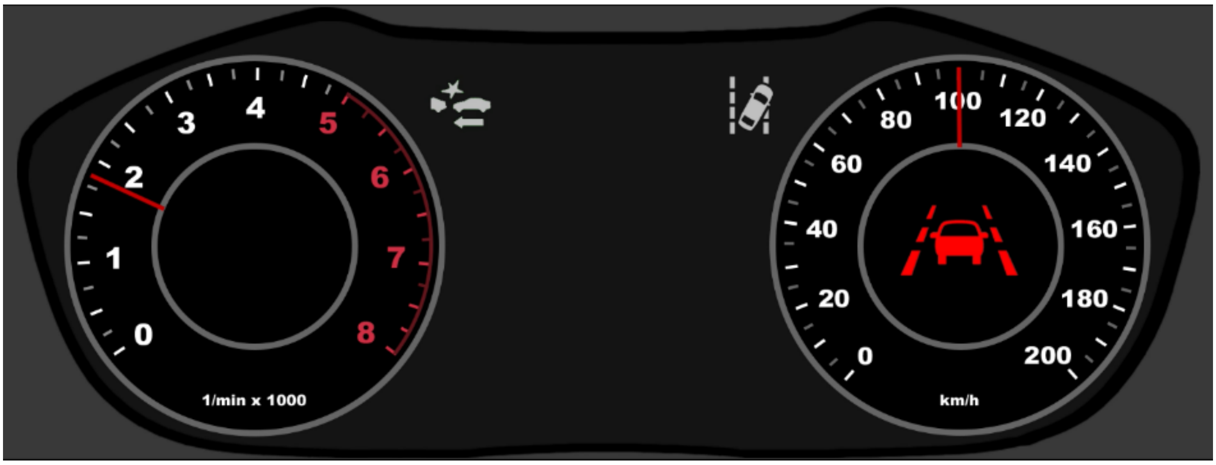


Transition





End



A.2 Audio

In all transition scenarios, an audio notification was played in the transition phase as part of the HMI. This consisted of repeated 1-second 'pling'- sound (containing multiple frequencies, fading out), with 1.5 s silence in between the sounds.

A.3 Steering wheel LEDs

The scenarios in which coloured LEDs on the steering wheel were used are listed in Table 5.1. The LED segments were activated on the steering wheel rim between 9 and 11 o'clock and between 1 and 3 o'clock. The LEDs were only activated in the transition phase. While activated, the LEDs were blinking in a pattern of 500 ms on, 500 ms off.

Table 5.1: Steering wheel LED colours per scenario

Scenario	Colour
17	Blue

18	Blue
19	Red
20	Blue
21	Blue
26	Red

Appendix B: Instructions for Participants

Wat fijn dat u vandaag mee wilt doen in ons experiment, genaamd "Onderzoek naar duidelijkheid en eenduidigheid van interfaces in auto's". Als het goed is hebt u al de informatiebrief over het experiment gelezen. In dit document geven we u extra uitleg over het experiment. Vragen kunt u altijd stellen aan de aanwezige proefleider van TNO.

Doel van het onderzoek

Auto's kunnen steeds meer zelf. Ondersteuningssystemen, systemen die de rijtaak gedeeltelijk overnemen, zoals Adaptive Cruise Control (ACC) en Lane Keep Assist bestaan al langere tijd. Er komen nu ook nieuwe automatiseringssystemen op de markt. Deze systemen nemen het rijden compleet over van de bestuurder, zodat die andere dingen kan gaan doen tijdens het rijden.

In deze studie onderzoeken we hoe we kunnen meten hoe duidelijk de informatie is die een auto over deze systemen geeft, bijvoorbeeld wanneer ze in- of uitgeschakeld worden en of ze een waarschuwing geven.

Het experiment

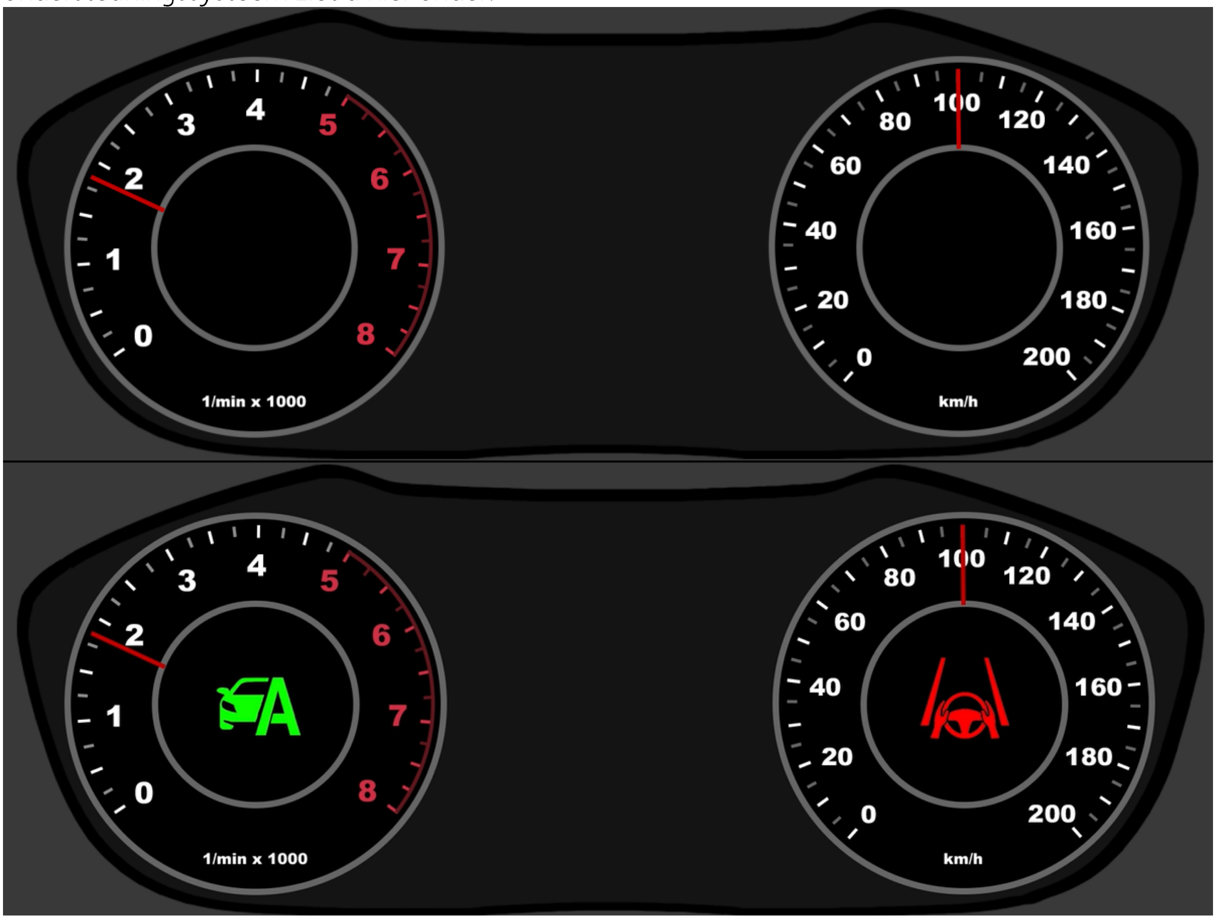
In dit experiment maken we gebruik van de rijsimulator. U zit daarbij in de cabine van een auto die op een bewegings-platform staat. U zit achter het stuur maar u gaat niet zelf rijden. We spelen een aantal "videofragmenten" af waarbij u op het scherm voor de simulator ziet hoe de auto over een snelweg rijdt. Vanaf de start van de rit rijdt de auto op constante snelheid, maar de rijsimulator zal hier niet bij bewegen. Tegelijkertijd kan er informatie voor u zijn van een automatiserings- of ondersteuningssysteem, in de vorm van symbolen of tekst op het dashboard en/of geluiden. Deze informatie die u krijgt (in beeld of geluid) noemen we "de HMI", wat staat voor "Human-Machine Interface".
Bij sommige filmpjes hoort u een computerstem, die zegt wat de bestuurder in deze situatie doet, bijvoorbeeld "Je besluit het stuur vast te pakken". Deze stem is alleen om toe te lichten wat er gebeurt tijdens het fragment en deze hoort niet bij de HMI. Hier laten we u straks een voorbeeld van horen.
De fragmenten duren ongeveer een halve minuut. Sommige filmpjes worden tweemaal afgespeeld. Nogmaals: u kijkt alleen naar deze filmpjes en u hoeft dus niet zelf de auto te besturen of systemen te bedienen. Na afloop van elk (herhaalde) videofragment vult u een korte vragenlijst in over wat u vond het de HMI. De knop om naar de vragenlijst toe te gaan verschijnt vanzelf op de tablet naast u. Eerst krijgt u een paar videofragmenten te zien om dit te oefenen. Daarna krijgt u in totaal 24 fragmenten te zien.

Procedure

1. Deze instructies doornemen
2. Computer-stem voorbeeld luisteren
3. Toestemmingsverklaring ondertekenen
4. Algemene vragenlijst invullen
5. Oefenfragment met vragenlijst
6. 24 video fragmenten met vragenlijst doorlopen
7. Afsluitend interview
8. Betalingsformulier invullen

Instructies

U krijgt straks verschillende HMIs en videofragmenten te zien in de rijsimulator. Een voorbeeld van een leeg dashboard en een met geactiveerde automatiserings- of ondersteuningssysteem ziet u hier onder.



In elk fragment is een automatiserings- of ondersteuningssysteem ingeschakeld óf, uitgeschakeld of wordt deze tijdens het videofragment in- óf uitgeschakeld. Na elk fragment krijgt u een aantal vragen over wat u zojuist gezien hebt om te bepalen hoe duidelijk de informatie was die de auto u gaf. Het is daarbij bedoeling om te letten op de automatiserings- of ondersteuningssystemen en om het hele HMI te beoordelen, dus inclusief eventuele geluiden. De vragen gaan niet over de snelheidsmeter en toerenteller; deze zullen in elke rit hetzelfde zijn. U krijgt vragen over:

- Wat denkt u dat er in de situatie aan het einde van het filmpje van u als bestuurder verwacht wordt?
- Een aantal vragen over hoe duidelijk u de informatie van de auto vindt.

Er zijn geen goede of foute antwoorden, het gaat ons er vooral om te bepalen of we met deze vragen kunnen meten hoe duidelijk de informatie van de auto is.
Na afloop van alle fragmenten stellen we u nog een aantal vragen over wat u is opgevallen aan deze HMIs en welke dingen een HMI duidelijk of onduidelijk maakten. Dit doen we in een afsluitend interview.

Vragenlijsten op de tablet

Onderstaande vragen kunt u verwachten na afloop van het filmpje. Indien u hier vragen over heeft, kunt u deze stellen aan de proefleider.

We willen graag weten hoe u de informatie in de auto tijdens het filmpje heeft begrepen. De volgende vragen gaan over de situatie aan het **eind van het filmpje** (die nu nog zichtbaar is) en wat u denkt dat u moet doen.

Wie moet er in deze toestand gas geven en remmen?

- ○ Ik doe dit helemaal zelf
- ○ De auto, maar ik moet continu kunnen ingrijpen
- ○ De auto doet dit zelf
- ○ Weet ik niet

Wie moet er in deze toestand sturen?

- ○ Ik doe dit helemaal zelf
- ○ De auto, maar ik moet continu kunnen ingrijpen
- ○ De auto doet dit zelf
- ○ Weet ik niet

Bent u verplicht om in deze toestand de omgeving in de gaten te houden tijdens het rijden?

- ○ Ja
- ○ Nee
- ○ Weet ik niet

Moet u in deze toestand uw handen aan het stuur houden?

- ○ Ja
- ○ Nee
- ○ Weet ik niet

Indien er een overgang plaats vond in het filmpje, dan krijgt u daarnaast nog de volgende vraag:

De volgende vraag gaat over het **hele filmpje** en wat er tijdens het filmpje veranderde.

In welke van deze dingen vond er verandering plaats in het filmpje?

- ○ Gas geven / Remmen
- ○ Sturen
- ○ Omgeving in de gaten houden
- ○ Handen aan het stuur houden
- ○ Weet ik niet

	Sterk mee oneens	Oneens	Enigzins oneens	Niet eens / niet oneens	Enigzins mee eens	Eens	Sterk mee eens
De informatie is dubbelzinnig	◯	◯	◯	◯	◯	◯	◯
De informatie kan meerdere betekenissen hebben	◯	◯	◯	◯	◯	◯	◯
Ik weet zeker wat deze informatie betekent	◯	◯	◯	◯	◯	◯	◯
De informatie is verwarrend	◯	◯	◯	◯	◯	◯	◯
De informatie is duidelijk	◯	◯	◯	◯	◯	◯	◯
Ik begrijp deze informatie snel	◯	◯	◯	◯	◯	◯	◯
De informatie is slecht leesbaar	◯	◯	◯	◯	◯	◯	◯
De informatie is moeilijk waar te nemen	◯	◯	◯	◯	◯	◯	◯
De informatie wordt duidelijk gepresenteerd	◯	◯	◯	◯	◯	◯	◯

Appendix C: Statistical packages used

R

- R version 4.5.1 (R Core Team, 2025)

LMM, ANOVA, marginal means, planned comparisons and post-hoc tests

- Lme4 package (Bates et al., 2015)
- emmeans package (Lenth, 2025)

Tests on proportions

- prop.test from the R Stats Package
- pairwise_prop_test from the rstatix package (Kassambara, 2023).

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Kassambara A (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. doi:10.32614/CRAN.package.rstatix <https://doi.org/10.32614/CRAN.package.rstatix>, R package version 0.7.2, <https://CRAN.R-project.org/package=rstatix>.

Lenth R (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. doi:10.32614/CRAN.package.emmeans <https://doi.org/10.32614/CRAN.package.emmeans>, R package version 1.11.1, <https://CRAN.R-project.org/package=emmeans>.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Signature

TNO › Mobility & Built Environment › Helmond, 14 November 2025

(Approved over email)

Bastiaan Krosse
Research Manager – TNO IVS

Jeroen Hogema
Primary Author

Mobility & Built Environment

Automotive Campus 30
5708 JZ Helmond
www.tno.nl

innovation
for life