

Final report

Development of a measurement scale to assess Automated Driving System HMIs

TNO 2023 R11506 – 07-08-2023

Development of a measurement scale to assess Automated Driving System HMIs

Final report

Author(s)	Jan Souman Ruben Post Jeroen Hogema Marika Hoedemaeker
Classification report	TNO Publiek
Report text	TNO Publiek
Number of pages	66 (excl. front and back cover)
Sponsor	Netherlands Ministry of Infrastructure and Water Management
Project name	User Centered Design of ADS HMIs
Project number	060.55112
Report number	TNO 2023 R11506

All rights reserved

No part of this publication may be reproduced and/or published by print, photoprint, microfilm or any other means without the previous written consent of TNO.

© 2023 TNO

Acknowledgments

The authors would like to thank Rutger Verstegen for contributing to this project in several ways as part of his M.Sc. project (HMI stimulus design, participant recruitment and test execution), Anja Langefeld en Robin Liefink for data collection, Wouda Visser-Oomen for participant recruitment and Frank Benders for reviewing the draft version of this report.

Summary

While the introduction of Automated Driving Systems (ADS) is intended to make driving safer and more comfortable, it may also make the interaction between user and vehicle more complex and prone to confusion. Users will sometimes be the active driver and at other times function as supervisor of the automation system. These different roles come with different responsibilities for the user in terms of vehicle control and traffic monitoring. This may create unintended safety risks in case the user is not aware of what his/her responsibilities are and what the capabilities of the automation system are. It is therefore important to be able to assess to what extent ADS provide their users with clear and unambiguous information concerning system state and concerning allowed or expected actions from the side of the user. In this project, TNO has developed a questionnaire to measure whether the information provided by an ADS to its user is experienced as clear and unambiguous. The questionnaire consists of nine items, which were shown to discriminate between good and bad Human Machine Interface (HMI) designs in terms of user perception and comprehension.

The development and evaluation of this questionnaire was done in four steps. First, a summary was compiled of existing methods and knowledge on the evaluation of HMIs with respect to “clear and unambiguous” and similar concepts. This was done based on a literature review and expert interviews, with a particular focus on automation systems. Knowledge from other relevant domains, such as automation in the medical and military domains as well as in aviation, was also included. The results showed that, while design guidelines and principles are often fairly concrete and specific for the low level perceptual characteristics of HMIs (e.g., use of colour and symbol size), they quickly become more general and abstract when it comes to how well the user can comprehend the information provided by a system. Moreover, quantitative measures of HMI quality are rare and no clear criteria for HMI evaluation are available. Hence, the current state of affairs is that each individual HMI has to be assessed on an ad-hoc basis with methods that need to be defined and implemented specifically for that HMI and are often qualitative rather than quantitative.

The second step within the project was to start defining questionnaire items for a more general HMI evaluation method, which is not only applicable to HMIs for different ADS, but also for a variety of user groups and use contexts. Initial items were formulated based on the literature review and expert interviews and were then ranked by a group of HMI design experts. This resulted in a set of 15 questionnaire items that could be presented as Likert scale items on a 7 point scale.

In order to test the reliability and internal consistency of the questionnaire as well as explore underlying factors, an online survey was conducted as a third step. In total 99 participants evaluated six HMI displays for six different ADAS using the set of 15 questionnaire items defined in step 2. The HMIs were designed to vary in the degree to which they provided clear and unambiguous information. Data analysis showed that responses on some of the items were highly correlated. A subset of 9 items was found to be able to discriminate well between HMIs with different levels of clarity and (un)ambiguity. Exploratory factor analysis revealed two underlying factors, which could be interpreted to relate to perception and comprehension of HMI information.

In the fourth and final step, the 9-item questionnaire was tested in an experimental setup. 23 Participants were presented with movie clips of transitions from or to automated driving, including a view of the traffic situation (through the windshield) and of the instrument cluster and steering wheel. Again, HMIs were designed to be either clear and unambiguous or unclear/ambiguous. The results showed that the questionnaire discriminated reliably between both versions of the HMI. As in the online survey, high reliability and internal consistency were observed. Additional questions intended to measure user comprehension more objectively turned out to be less useful. In contrast, qualitative measurement of user experience by means of Product Reaction Cards also showed clear differences between the two different HMI versions.

While the questionnaire discriminated successfully between clear and unclear HMIs, it only measured user perception and comprehension, not user response, as this was beyond the scope of the project. Obviously, in real life user response is relevant, as this determines the safety impact of an HMI. Therefore, further research should focus on the extent to which an HMI promotes adequate and timely user responses. Several other relevant factors, such as user diversity, user attention and traffic situation should be considered as well. Moreover, in order to use the questionnaire in HMI assessment, a reference point should be established by means of a benchmark HMI.

Contents

Summary	3
Contents	5
1 Introduction	7
2 Existing approaches	9
2.1 Literature scan	9
2.2 Literature overview regarding HMI evaluation in different domains.....	9
2.2.1 Automotive	9
2.2.2 Aviation	11
2.2.3 Medical	11
2.2.4 Military.....	12
2.3 Subject Matter Experts interviews	13
2.3.1 Euro NCAP.....	14
2.3.2 Automotive design researcher	15
2.3.3 Interaction designers.....	15
2.4 Summary of literature scan and expert sessions.....	16
3 Definition and operationalisation	20
3.1 Defining Clear and Unambiguous for Automotive HMIs.....	20
3.1.1 Definition of 'Clear and Unambiguous'	20
3.1.2 Relation to safety.....	20
3.2 Measuring Clear and Unambiguous for Automotive HMIs.....	21
3.3 Item development	22
4 Survey Study	24
4.1 Introduction.....	24
4.2 Method.....	24
4.2.1 Participants	24
4.2.2 Stimuli	24
4.2.3 Procedure	27
4.3 Results	28
4.3.1 Manipulation check	28
4.3.2 Exploratory Factor Analysis	30
4.4 Discussion.....	33
5 Experimental tests and results.....	34
5.1 Introduction.....	34
5.2 Methods	34
5.2.1 Participants	34
5.2.2 Stimuli	35
5.2.3 Measures.....	36
5.2.4 Design and procedure.....	38
5.2.5 Analysis	39
5.3 Results	40
5.3.1 Experience with ADAS and affinity for technology	40
5.3.2 Open questions results.....	41
5.3.3 Responsibility questions	42

5.3.4	Clear and Unambiguous Questionnaire	42
5.3.5	Product Reaction Cards	44
5.4	Discussion	45
6	Conclusions and recommendations	47
6.1	Summary	47
6.2	Discussion	48
6.2.1	User (groups)	48
6.2.2	Towards task performance	48
6.2.3	Information to users	49
6.2.4	Relative rating of clear and unambiguous	49
6.3	Recommendations	49
7	References	51
8	Abbreviations	53

1 Introduction

With increasing deployment and further development of Advanced Driver Assistance System (ADAS) as well as Automated Driving Systems (ADS), car driving is being increasingly automated. For the time being as well as in the foreseeable near future, this concerns mainly conditional automation (SAE, J3016. The Principles of Operation Framework: A Comprehensive Classification Concept for Automated Driving Functions., 2021). Automation is only available in restricted Operational Design Domains, such as on highways when following another vehicle at low speed with good visibility. This implies that frequent transitions of control from human driver to ADS and vice versa will occur. To ensure safe transitions of control from one to the other and to avoid mode confusion, good quality of interaction between user and system is required. In this context, there is a need for test procedures with associated criteria by which the quality of interaction between user and vehicle systems can be assessed. In order to achieve safe and effective interaction, it seems obvious to adopt a User Centred Design (UCD) process (de Goede, Jansen, & van Grondelle, 2023). Documentation created in this process by an OEM can also be part of audits that may become part of future vehicle certification. Requirements that the system must meet are currently available only at a high and abstract level. 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.

Recently, SWOV and TU Delft have conducted a study in which they interviewed OEMs on how they apply user centred design (de Goede, Jansen, & van Grondelle, 2023). This resulted in a description of how the design process is currently implemented by industry, and how it can be improved and assessed. The work in this TNO report focuses not on the design process, but on how an HMI resulting from such a design process can be tested in terms of how clear and unambiguous the information provided by the HMI to the user is.

The report is set up as follows:

Chapter 2 presents the results of a literature scan of existing theoretical knowledge of what it means for an HMI to provide the user with 'clear and unambiguous' information. Not only guidelines and standards from the automotive domain, but also from aviation, the military and the medical domains are considered. In addition, we have interviewed a small number of subject matter experts to add their practical knowledge.

In Chapter 3, the definition and operationalisation of "Clear and Unambiguous" design is presented. This definition is based on the findings in Chapter 2. Furthermore, questionnaire items were generated that could measure how "Clear and Unambiguous" an HMI is.

In Chapter 4, a survey study is described that assesses the questionnaire items. The list of items is reduced based on the study outcomes to create a scale that measures "Clear and Unambiguous".

In Chapter 5 an experimental study is presented in which a set of systematically manipulated HMIs is used to re-test the scale for reliability in a more controlled and dynamic situation and to assess additional measures for “Clear and Unambiguous”. Chapter 6 gives the conclusions and recommendations on if and how “Clear and Unambiguous” can be taken up in the testing and certification process of new vehicles.

2 Existing approaches

As input for the definition and operationalisation of what it means for an HMI to provide the user with ‘clear and unambiguous’ information, existing theoretical and practical knowledge was collected in two ways. First, the existing scientific literature and design standards were scanned for relevant information. Second, interviews were conducted with a number of subject matter experts, both internal and external to TNO. Both activities are described below in more detail. Their joint results are combined in Section 2.4 to provide a summary and initial conclusions.

2.1 Literature scan

A literature scan was performed to assess the availability and content of scientific and otherwise relevant literature such as design standards and guidelines on definitions and measures of clear and unambiguous (C&U), or associated concepts, related to information presentation in HMIs. The scan focussed on the automotive domain, and was complemented with literature from the aviation, medical and military domains. These domains were added due to their commonalities in designing and evaluating complex HMIs for safety critical and user operated systems that involve automation. Furthermore, we restricted the search to the perception (e.g., legibility, audibility, contrast) and cognition (e.g., comprehension, understanding, interpretation) phases of information processing (Raab, Lobinger, Hoffmann, Pizzera, & Laborde, 2015), as the acting phase (e.g., planning and deciding on a course of action) is out of scope for the current research. Lastly, while HMI design is heavily focussed on the visual modality, auditory and tactile modalities were included where appropriate.

The selection of documents to include was based on the content at least describing one of the following:

- Guidelines, guidance, criteria, considerations or industry best practices for HMI information presentation
- A definition of clear and unambiguous
- Subjective or objective measures of clear and unambiguous
- Methods to assess clear and unambiguous

Literature search was performed using Scopus, Google search and Google Scholar as well as by citation tracking of relevant publications. The resulting documents were stored in a digital repository and a subset of the most relevant documentation for further analysis was made through team discussions.

2.2 Literature overview regarding HMI evaluation in different domains

2.2.1 Automotive

The European Statement of Principles on HMI (ESoP; European Commission, 2008) provides a long list of principles for the design of information and communication systems, which also

apply to ADS HMIs. Of these, the so-called information presentation principles are most relevant for the evaluation of the extent to which information provided by a HMI is clear and unambiguous:

- Visually displayed information presented at any one time by the system should be designed in such a way that the driver is able to assimilate the relevant information with a few glances which are brief enough not to adversely affect driving.
- Internationally and/or nationally agreed standards relating to legibility, audibility, icons, symbols, words, acronyms and/or abbreviations should be used.
- Information relevant to the driving task should be accurate and provided in a timely manner.
- Information with higher safety relevance should be given higher priority.
- System-generated sounds, with sound levels that cannot be controlled by the driver, should not mask audible warnings from within the vehicle or the outside.

The ESoP specifies to which type of information these principles apply and provides examples of good and bad implementation. It also provides a verification method for each principle. However, these are formulated in a very general and abstract way, without clear operationalisation.

According to Campbell et al. (2018), comprehension of messages from the vehicle HMI consists of three stages, all of which should be considered in HMI design and evaluation:

- Extraction: Can the driver see/hear/feel the message fully and accurately under a representative range of driving circumstances and conditions? This relates to the perceptibility of the message: visual size of symbols or characters, contrast, colour, labels, etc. for visual messages; sound level, loudness, pitch, etc. for auditory messages; for haptic messages: type, location, frequency, amplitude etc.
- Recognition: How well do parts of the message relate to each other? Does the construction of the message support accurate understanding? Can the message be easily confused with other messages? This relates to the temporal characteristics of a message, as well as to the level of realism and detail.
- Interpretation: How well does the message reflect its underlying meaning? Will it be understood when presented in the appropriate context? Does it require any special knowledge, particular to a culture, language or driver age? This relates to the use of colour, cues for the relative urgency, for external locations and for multimodal cues.

Campbell et al. (2004) proposed and tested a method to evaluate the comprehension of static HMI symbols or icons for active safety systems. This method has been incorporated into the SAE standard J2830 (SAE, 2016), which generalizes it to all static HMI symbols. Campbell et al. (2004) asked participants for various icons to write down what they thought the icon meant, as well as to rank a set of icons according to how well they represented a certain message. Participant interpretations of icons were categorized into nine response categories, including a separate category for safety critical confusions.

Naujoks et al. (2019a; 2019b) suggested a checklist approach for the evaluation of ADS HMIs. The 20 checklist items evaluate extraction, recognition and interpretation of information as defined by Campbell et al. (2018), as well as the resulting action (especially preventing unintentional activation or deactivation of an ADS). The authors do not provide a metric to score the degree to which this checklist is met, but they give minimum criteria for acceptable versus bad HMIs. The checklist should be used by experts in combination with testing with user groups.

Recently, Liu et al. (2022) suggested a measure for the understandability of a HMI, which they called 'functional transparency'. It is based on the subjective evaluation of the user

whether they understand an HMI and on their answers to a set of probe questions. The functional transparency measure is defined as:

$$T_{functional} = \begin{cases} 0, & \text{if the user reports not to understand the HMI} \\ AU \left(1 - \frac{TNPU}{TNPU_{max}}\right), & \text{otherwise} \end{cases}$$

where AU represents the actual understandability (percentage correct answers), $TNPU$ is the time needed for perceived understandability (in s), and $TNPU_{max}$ is the maximum allowed time to perceive understandability (as a normalizing constant). Thus, functional transparency is highest when a user can correctly answer all probe questions in the shortest amount of time. Note that there is no lower threshold on response time $TNPU$. Consequently, functional transparency will always be lower than 1.

2.2.2 Aviation

The aviation industry has thorough documentation on the design considerations and guidance for aircraft cockpit design (European Aviation Safety Agent, 2018; (Yeh, Swider, Jin Jo, & Donovan, 2016). These documents include sections on the need for clear and unambiguous information presentation (Certification Specification CS 25.1302 and Chapter 3.1, respectively).

According to CS 25.1302, the clear and unambiguous design of flight deck systems implies that its information:

- Can be perceived correctly (is legible).
- Can be comprehended in the context of the flight crew task.
- Supports the flight crew's ability to carry out the action intended to perform the tasks.

Besides a description of these different aspects of HMI information, the document provides design guidance and recommendations for creating clear and unambiguous flight deck systems. Although both documents include examples of factors that influence how clear and unambiguous an HMI is, measuring if these factors were successfully applied to create clear and unambiguous flight deck systems (i.e., compliance) is not described in detail. Instead, CS 25.1302 states that this “should be done on a case-by-case basis, driven by the specific compliance issues” (pp. 674). These should be developed and proposed by the applicant, and then agreed to by the Agency (i.e., EASA). Means of compliance in the form of various tests are presented to aid the applicant in determining suitability. Additional documentation on alarm design provides input on auditory factors (Beaujard, 2018).

Even though at first glance the task of a pilot may resemble that of a driver in several respects, there are important differences that should be taken into account when applying design principles from the aviation domain in an automotive context. Most importantly, driving a car typically puts much shorter time limits on vehicle control than flying an aircraft. The HMIs of all but the smallest aircraft are also far more complex than those in vehicles. Moreover, pilots are highly trained professionals, who are required to update their knowledge and skills regularly. Instead, knowledge, skills and experience vary widely among car drivers. These factors all put different demands on vehicle HMIs than aviation HMIs.

2.2.3 Medical

Medical devices and their interfaces are designed to be used safely and effectively by following a Human Factors Engineering (HFE) process (US Food and Drug Administration,

2016). This process provides both guidance on how to design safe and effective interfaces, as well as how to validate the design's ability to achieve this. Central to this process is taking a user-centred design approach. This implies applying design standards and recommended practises during interface design (Association for the Advancement of Medical Instrumentation, 2018) and iterative testing of their impact on safety and effectiveness.

While the terminology “clear and unambiguous” is sparsely used in these documents, a safe and effective medical interface generally implies clear and effective information presentation: “Controls, displays, and other equipment features that need to be located, identified, or manipulated should be appropriately and clearly marked to permit rapid and accurate human performance” (pp. 139). Such can be supported by adhering to the following interface requirements:

- Functional elements that are not intuitively obvious and require identification should be labelled or marked.
- Markings should be positioned so that they are clearly associated with the correct equipment feature and not obscured by hand positions or equipment components.
- Appropriate markings can enhance the identification of both individual elements and their functional relationships.
- Ambiguous symbols, codes, or terminology should be avoided.
- Designing for legibility requires careful analysis of ambient illumination in typical use environments.
- Consistency of placement, terminology, and coding is critical.
- All markings should be tested with typical users. Users can be clinicians, caregivers, patients, or maintenance personnel and can vary by age, disability, and other characteristics.

Besides the above mentioned requirements, the HFE process also provides insight on how to measure if a medical interface can be used safely and effectively, which can inform measures of clear and unambiguous. In particular, the FDA (US Food and Drug Administration, 2016) states that:

The testing should demonstrate that:

1. the device can be used by the intended users without serious use errors or problems.
- The testing should be:
2. for the intended uses and
 3. under the expected use conditions.

The human factors validation testing should be designed as follows:

- The test participants represent the intended (actual) users of the device
- All critical tasks are performed during the test
- Sample size is best determined from the results of preliminary analyses and evaluations.
- Manufacturers should make their own determinations of the necessary number of test participants but, in general, the minimum number of participants should be 15 for each user population.

2.2.4 Military

Similar to the medical domain, the interfaces developed for military use adhere to design criteria focussing on the users' ability to quickly and correctly assess its presented information. That said, the military domain also poses unique human factors issues compared to automotive. For example, military users might wear (facial) equipment limiting their vision.

They may also be under higher cognitive load compared to regular drivers. Furthermore, they are likely more trained for the specific system they are working with. These and other factors prevent directly applying design criteria to the automotive field. That said, documentation such as MIL-STD-1472H (Department of Defense, 2020) does provide direction on design features influencing clear and unambiguous information presentation in the areas of:

- General visual display requirements (e.g., Orientation of objects, Flash coding)
- Electronic display requirements (e.g., Contrast ratio, Legibility)
- Scale requirements (e.g., Linear scales, scale numerals)
- Automation (e.g., Indication of operating mode, Functional use of color)
- Speech and Audio Systems (e.g., Signal meaning, Verbal warning signals, Comprehension)
- Labelling (e.g., Location, Orientation, Visibility and legibility)
- Warnings (e.g., Hazards and Signal words)
- Information Systems (e.g., Notifications, User guidance, Information content, Scope, Formatting, Organization)

As stated before, the application of such military design criteria to support clear and unambiguous automotive HMIs requires assessing their relevance given the user, task and use environment, and translating them where appropriate. Whereas some design criteria are generalizable across domains (e.g., each audio signal shall have only one meaning), others are not. As an example, a design criterion for visual displays states that the orientation of displayed objects shall include reference to the vertical or horizontal direction. Such information might not be particularly informative, or even applicable, for automotive HMIs.

2.3 Subject Matter Experts interviews

Two interviews (each about 1 h) were conducted with experts in the field to verify that we had collected the most important information concerning what it means for an HMI to provide the user with ‘clear and unambiguous’ information. One of the experts had a background in vehicle safety assessment, while the other one was an academic with experience in automotive design. A third expert from the automotive industry was invited, but her calendar did not permit an interview to be scheduled within the timeframe of this project. In addition to the two interviews, TNO internal interaction designers were asked for input on how to design HMIs that are clear and unambiguous.

The interviews were semi-structured. First, the goal, background and intermediate results of the project were presented to the experts, using Powerpoint slides. Then, the experts’ opinions were asked in a free format conversation, in which at least the following questions were addressed:

- Do you recognize the challenge of evaluating human-machine interaction with automated driving systems? What are your thoughts on how to deal with this?
- Which elements of our definition of ‘clear and unambiguous’ do you recognize? What do you miss?
- How do you think ADS HMIs could be assessed?
- What are your thoughts about the feasibility of our suggested approach (C&U Questionnaire)?
- Current guidelines largely focus on perceptual aspects of HMIs. How can design help to predict/assess comprehension of information?
- How should the user response in terms of action/behaviour be taken into account in HMI assessment?

2.3.1 Euro NCAP

Euro NCAP is an independent non-profit organization that tests car safety using a five-star safety rating system. Founded in 1997, Euro NCAP has more or less become the standard for indicating how safe a car is, both in terms of crash prevention and of the impact of crashes in case they cannot be prevented. The five-star safety rating system continuously evolves as older technology matures and new innovations become available. This means that tests are updated regularly, new tests are added to the system and star levels adjusted. Euro NCAP has also started to include HMI evaluation in their safety assessment protocols. Because Euro NCAP has ample experience in the development and execution of vehicle test protocols, we had a discussion on the topic of clear and unambiguous HMI with Richard Schram, technical director of Euro NCAP.

In the view of Euro NCAP, there is no transition of control in assisted driving (up to and including SAE L2). Currently, real automated driving is not (yet) available in the Netherlands. Hence, the driver is always in control and responsible for all aspects of driving. In that sense, there is no risk of mode confusion. In automated driving, transition of control can take place. In designing these automated systems, priority should be given to ensuring safe transition of control, preferably also in a comfortable way. For this, it is important that the driver understands the request to take over control, which means that it should be clear and unambiguous. The system must have a solution for all situations where a driver does not respond to the take-over request, but this solution does not have to be comfortable (like automatic emergency braking or manoeuvring the car to the emergency lane). For the user, it may be unclear what the difference between assisted and automated driving is. Consequently, Automated Lane Keeping System (ALKS) regulation 157 (United Nations, 2023) states that on deactivation of the ALKS the vehicle must revert to manual driving, not to assisted driving.

Instead of trying to assess whether an HMI provides clear and unambiguous information, it might be easier to determine whether the information is unclear or ambiguous, thus excluding unsafe HMIs. However, while this might be advantageous from an assessment perspective, it does not provide designers with clear guidelines with respect to ADS HMIs. Euro NCAP is against testing or surveys with groups of naive users. The HMI design of a vehicle may be very clear and unambiguous to users accustomed to the design of that brand, while it can be less clear to new users. OEMs place great value on brand specific interior designs. One possibility to incorporate the feedback from both experienced and unexperienced users into HMI evaluation might be to attribute different weights to different user groups in the assessment procedure.

Euro NCAP has a checklist to measure driver engagement in assisted driving. However, currently this checklist is still quite general and needs to be further developed. A good HMI test procedure would be a welcome addition. Apart from the displays that provide the user with information, it is important to also take a vehicle's controls into account when assessing an HMI. They play an important role in the user's interaction with the vehicle and consequently are important for the user's mental model of the vehicle in general and assistance/automation systems in particular. Moreover, the assessment of how clear and unambiguous an HMI is might differ for information versus warning or intervention by a vehicle system.

2.3.2 Automotive design researcher

Elmer van Grondelle worked many years as a car designer for a number of OEMs. Currently, he teaches automotive design and mobility design at the Technical University of Delft (TU Delft) and works with students on designing automotive interfaces. For example, within the EU project Mediator he was responsible for setting up the design guidelines that were validated in driving simulators and real vehicles of several OEM partners. In his work, Van Grondelle focuses on a systems approach of design management, looking not only at HMI design, but also taking into account the societal developments towards a changing role of the vehicle in mobility.

According to Van Grondelle, autonomous cars without a steering wheel are not expected for at least 40-50 years. Consequently, transition of control is an important topic. There is no proof that User Centered Design leads to better HMIs. Nevertheless, it is important to take basic design principles into account:

- Design for existing affordances (for example, red means danger, gear stick has always same form suggesting the possibility to move it to different positions).
- For interfaces of functions that don't exist yet, users often need prior knowledge to understand and use it.
- Experienced users want less information than inexperienced users, so adaptability of the information is important.
- Information should be unambiguous.
- Information should be clear for first time users, for instance when people change brands or use a rental car. This should be part of HMI assessment.

Timeliness is an important part of the definition of 'Clear and Unambiguous'. It is, however, also a part that is more difficult to test. In HMI assessment, it is important to consider the time users will have to evaluate or use an HMI. Having unlimited time may lead to different results (e.g., different interpretation) than having to decide in a short amount of time. In order to test the comprehension of HMI information, tests with users are necessary (expert opinion may not suffice).

Designers in OEM design studios are primarily concerned with brand experience. In the future, brand experience will only become more important to differentiate a brand from the competition. As a consequence, the focus is on user experience, less on user comprehension. It would be interesting to test current car HMIs against the clear and unambiguous criteria to see how well they fit to these requirements.

In setting criteria for ADS HMIs, we need to look at what is a reasonable expectation regarding the safety for automated cars. In the assessment of current vehicles, we do not demand absolute safety. Similarly, the autonomous car will not be perfect and we need to accept a certain level of unsafety.

2.3.3 Interaction designers

TNO has a dedicated Human Machine Teaming department at TNO Soesterberg, including a team of interaction designers and design researchers. They have experience supporting the design of HMIs for a wide range of domains and systems, including the automotive domain. A 30-minute online session was organized with eight participants from this team. Participants were given a brief introduction on this project's research question and asked to provide design recommendations and principles that support "Clear" and "Unambiguous" HMI design. Answers were logged in an online whiteboard and analysed by the research

team. Similar answers were grouped together and additional clarification on suggested design recommendations and principles were asked in follow-up sessions.

The resulting design recommendations and principles were:

- Conciseness
- Standardization of terminology
- Actionable terminology
- Make information easy to process
- Take context and user into account
- Create (visual) hierarchy
- Avoid glare, low contrast, and low brightness
- Provide meaning
- Relate information to context
- Present information to the most suitable sensory modality
- Provide mode indications

2.4 Summary of literature scan and expert sessions

The previous sections provide a brief overview of the vast amount of information available in the scientific literature as well as in design guidelines and standards with respect to HMI design and evaluation. [Table 2.1](#) provides an overview of the most relevant guideline documents for HMI design and evaluation. For a more exhaustive list of relevant standards, we refer to Campbell et al. (2018).

While a comprehensive review was outside the scope of the current project, the literature scan combined with our interviews with subject matter experts suggests that as yet there is no consensus on how to evaluate in a systematic and quantitative fashion whether an HMI provides clear and unambiguous information beyond the low-level (perceptual) aspects such as visibility, contrast, legibility, audibility, etc. These perceptual factors are however crucial for correct comprehension, and as such, it is essential that the user correctly perceives an HMI. We identified a list of conditions that are necessary for an HMI to support it being clear and unambiguous. That said, adherence to these conditions does not automatically imply the resulting HMI is clear and unambiguous. They define necessary, but not sufficient conditions for an HMI to be clear and unambiguous.

The conditions that we identified can be grouped into several categories.

General

- Use actionable terminology
- Make information easy to process
- Take context and user into account
- Create (visual) hierarchy
- Provide meaning
- Relate information to context
- Present information to the most suitable sensory modality
- Provide mode indications
- Design for existing affordances of drivers
- Avoid drawing the driver's attention away unnecessarily from impending danger

Readability

- Consider characters, font, lines and (scale) markings
- Avoid glare, insufficient contrast and insufficient brightness
- Account for visual angles
- Account for adverse conditions, such as vibration
- Use intuitive abbreviations
- Use consistent and standardized terminology
- Avoid ambiguity of terms or phrases
- Create simple phrases
- Make use of proper indenting

Simplicity / Conciseness

- Aim for the simplest design that conveys the necessary information.
- Displays should not contain extraneous information, text, or graphics
- For text displays, the display density (area occupied by text) should be less than 50%; a display density of 25% or less is preferable.
- Information should be presented using the least precise display format

Colour

- Respect color-coding throughout the HMI
- Consider color-coding of other HMIs

Use and position of symbols and/or labels

- Functional elements that are not intuitively obvious and require identification should be labelled or marked
- Markings should be positioned so that they are clearly associated with the correct equipment feature and not obscured by hand positions or equipment components
- Appropriate markings can enhance the identification of both individual elements and their functional relationships
- Labels should be visible at typical viewing distances and angles
- Labels should be resistant to wear and tear
- Ambiguous symbols, codes, or terminology should be avoided

Arrangement and Organization

- Displays should be organized by function, sequence, access, and/or nominal viewing angle
- Information on a display should be organized by sequence, importance, function, frequency
- If displays or information on displays are arranged sequentially, the information should be viewed left to right when arranged horizontally or top to bottom when arranged vertically
- Sets of data that are associated with specific questions or related to particular functions shall be grouped together to signify those functional relationships
- All information needed for one task should be located on the same display
- The user should not need to remember information across pages

Adhering to the above perceptual factors supports the design of a clear and unambiguous HMI. However, while such factors can be measured on their own, and sometimes jointly, their interrelated nature prevent drawing conclusion about the HMI as a whole. A more holistic measure is needed to understand if the combined perceptual conditions together create a clear and unambiguous HMI.

Table 2.1: Overview of existing guidelines for HMI design and evaluation in different domains.

Title	Author	Year	Domain	Description
European Statement of Principles on safe and efficient in-vehicle information and communication systems	EC	2008	Automotive	Information presentation principles
Human Factors Design Guidance for Level 2 And Level 3 Automated Driving Concepts	Campbell et al.	2018	Automotive	Guidelines for the design of L2/3 HMIs
Human Factors Design Guidance For Driver-Vehicle Interfaces	Campbell et al.	2016	Automotive	Guidelines for HMI design
Human Factors Guidelines for Advanced Driver Assistance Systems and Automated Driving Systems	Souman et al.	2020	Automotive	Guidelines for the design and evaluation of L2-4 HMIs
Easy Access Rules for Large Aeroplanes (subsection: AMC 25.1302)	EASA	2018	Aviation	Design considerations and guidance
Overview of the HEAA method defined by Airbus for Alarm design (Human Errors Analysis which concentrates on Alarm titles and their procedures)	Beaujard	2018	Aviation	Criteria for clear and unambiguous alarm design
Human factors considerations in the design and evaluation of flight deck displays and controls	Yeh et al.	2016	Aviation	Guidance for the design and format of information elements
Human factors engineering – Design of medical devices (chapters 10 and 19)	AAMI	2009	Medical	Design guidelines and guidance for designing safe and effective medical devices
Applying Human Factors and Usability Engineering to Medical Devices	FDA	2016	Medical	Human Factors Engineering guidance for designing safe and effective medical devices
MIL-STD-1472H – Human Engineering	US MoD	2020	Military	General human engineering design criteria for military systems, subsystems, equipment, and facilities

Although no explicit definitions of clear and unambiguous were found, several aspects that should be considered for such a definition were mentioned in the literature and in the expert interviews. First, evaluation of an HMI should consist of more than just the low-level perception such as legibility or audibility. Comprehension and the consequences for driver action should be considered as well, as these ultimately determine the impact on safety. Second, not only the HMI under consideration, but also the (intended) user groups, the tasks for which an HMI will be used and the use conditions should be considered when evaluating if an HMI provides clear and unambiguous information. Third, ‘clear and unambiguous’

represent two distinct, though related, concepts. The question is whether they should be defined and measured together, or that they should be considered separately.

Although many guidelines for HMI design can be found in the literature, actual measures and criteria for the evaluation of HMIs against these guidelines are far more scarce and some authors are sceptical whether generally applicable criteria can be established at all (Carsten & Nilsson, 2001). Nevertheless, some different approaches to the evaluation of HMIs can be found. Similar to the common approach in aviation, Naujoks and colleagues (2019a; 2019b) have proposed a checklist to evaluate HMI design, in particular for automated driving systems. While the scope of this checklist is geared towards safety and thus broader than whether an HMI provides the user with clear and unambiguous information, it does not come with clear scoring metrics. In contrast, Liu et al. (2022) presented a quantifiable metric of functional transparency, which appears to largely overlap with 'clear and unambiguous'. However, they also do not provide a criterion for evaluating whether an HMI's functional transparency is good enough. Moreover, the metric depends on a set of user questions that are specific to the HMI under testing and hence not generally applicable.

In addition, the literature scan revealed several points that need to be taken into account when developing test procedures for HMI evaluation. In particular, the user group involved in testing should be representative of the intended user group, use conditions should include the variation encountered in day to day use of the system and the HMI should be tested with at least the most safety-critical tasks that it will be used for in expected use.

3 Definition and operationalisation

3.1 Defining Clear and Unambiguous for Automotive HMIs

The research team synthesized the results from the activities performed in Chapter 2 to develop a definition of Clear and Unambiguous.

3.1.1 Definition of ‘Clear and Unambiguous’

We define the information provided by an HMI to its user(s) to be clear and unambiguous when it allows the user to perceive and understand the information correctly and timely in all relevant use contexts.

Even though this definition applies to all kinds of HMIs, in this report we focus on automotive HMIs, specifically those relevant for safe and effective execution of the driving task, be it at the operational level of direct vehicle control, at the tactical level of manoeuvring the vehicle or at the strategic level of trip planning (including navigation) (Michon, 1985). The term ‘user’ can apply to all occupants of a vehicle, both drivers and passengers, including users with changing roles as in the case of transition of control. The role of HMIs in driving automation is particularly relevant, given the added complexity in driver-vehicle interaction. Relevant aspects of user diversity, for instance in terms of use experience, cultural/linguistic background, perceptual and cognitive abilities or age, must be considered in the evaluation of HMIs. In addition, the range of relevant use contexts should be determined. These include not only the driving mode of the vehicle (automated, assisted or manual driving), but also the current traffic situation and the demands it puts on the user. In addition, the user situation in the vehicle should also be considered (single user or multiple occupants, user physical or mental state, noise and distraction levels in the vehicle cabin).

This definition of ‘clear and unambiguous’ focuses on user comprehension, not on user action. Thus, even though the information should enable the user to choose and execute appropriate actions (or defer from action), the actual action is not part of whether information is ‘clear and unambiguous’. Taking appropriate action requires more than the availability of information that is clear and unambiguous. It also calls for the ability to weigh this information, to consider other information or the lack thereof, to take decisions based on the available information, to choose the appropriate action and to execute this action.

3.1.2 Relation to safety

HMI design and evaluation are important, not only because the HMI largely determines the look and feel of a vehicle, which makes it attractive to the intended user group, but also because it has consequences for the safety of the vehicle when used. This is especially true for ADS HMI and transitions of control. Without clear and unambiguous information from the vehicle to the user about its current state and about the expected behaviour of the user, changes in ADS behaviour (e.g., switching off when the ODD is exceeded) may lead to disastrous outcomes. However, it is important to realize that HMI design is only one factor amongst many that determine whether someone uses a system safely. As depicted in

Figure 3-1, HMI design (and thus how clear and unambiguous an HMI is) has a major impact on how well a user is able to perceive the HMI message. However, both its relative and absolute importance decrease when looking at user responses further along the processing continuum. In terms of actual action in response to HMI messages, other factors such as user state, user traits and the conditions under which the user receives the information (such as traffic, visibility, etc) also play a (possibly even bigger) role than HMI design. Thus, while clear and unambiguous HMIs are important and necessary for safe use, they are not sufficient and can never guarantee safety.

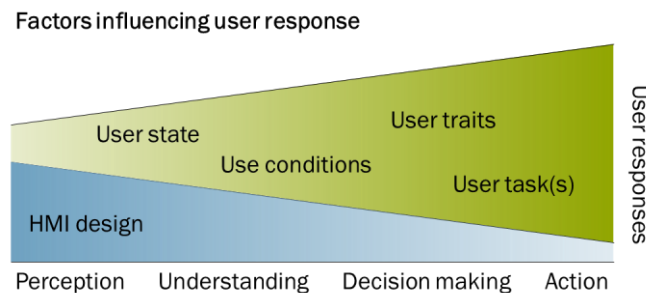


Figure 3-1: User response variability as a function of processing stage. The influence of HMI design is biggest for the perception of HMI messages, but its relative importance decreases towards actual action, where many other factors play a role.

Carsten and Nilsson (2001) suggested that standardized performance assessment of automotive HMIs to evaluate their safety is not generally possible for all types of systems. According to the authors, it should be possible to develop evaluation criteria for information systems, but not for systems that warn the driver or intervene in the driving task. With these types of systems, safety impact goes beyond the comprehension of information. Aspects such as underload of the driver, mode awareness/confusion and behavioural adaptation have to be taken into account as well. For this reason, the authors recommend a process-oriented approach for these types of systems.

3.2 Measuring Clear and Unambiguous for Automotive HMIs

Whether information is presented in a clear and unambiguous manner depends on the interaction between multiple components. User, use environment and the HMI are components that together influence how the presented information is perceived and comprehended. These components themselves can vary in many ways (e.g., different users, different contexts, different interfaces). How these components affect each other is currently not fully understood. Nonetheless, it is particularly important to operationalize clear and unambiguous in such a way that its measure is flexible enough to take into account this variety of users, for diverse (both current and future) use environments and for HMIs that may involve technologies that are not yet used. This combined complexity makes operationalizing clear and unambiguous in solely mathematical formulae currently unachievable. More appropriate are rating scales, as they have been successfully developed to measure abstract and subjective perceptual concepts such as system usability (Brooke, 1996), product aesthetics (Blijlevens, Thurgood, Hekkert, Chen, & Leder, 2017) and user satisfaction of HMIs (Chin, Diehl, & Norman, 1988). However, to our knowledge no such scale exists for clear and unambiguous. We therefore developed a scale with the aim to measure clear and unambiguous for ADS HMIs.

3.3 Item development

In an exploratory phase, the research team used the findings from the literature scan and expert sessions to generate a list of 24 questionnaire items (in English) that could measure the constructs of “Clear” and “Unambiguous”. These items were reworded into Likert scale-type items (see [Table 3.1](#)). In a second phase, the initial list of items were translated to Dutch, as they were intended to be presented to a Dutch participant sample. As such, results from the questionnaire development should only be considered in light of a Dutch population. Here, however, items are written in English for consistency in the report body. The Dutch questionnaire is documented in Appendix A. All items were ranked by seven experts using an online form. The form included a brief introduction on the purpose of the research and asked participants to rank items on their appropriateness and relevance in measuring whether an automotive HMI is clear and unambiguous. The resulting rank order of items was averaged across responses and standard deviations were calculated for each item. Items with exceptionally large standard deviations, indicating a lack of consensus between experts, were deleted. For items that were highly similar in meaning and/or phrasing, the lowest ranking items (i.e., worst scoring) were removed. Of the remaining items, the 15 highest ranking items were kept. Some of these used the same concepts, but were formulated either as a characteristic of the information or as a state of the user (e.g., Q1 and Q2; Q11 and Q13), to investigate which formulation was better. At this stage of development, these similarly looking items were allowed to explore potential differences in responses to variations in wording. The 15 items are presented in

[Table 3.2](#) and were used for the exploratory analysis in the Survey Study to further develop the scale.

Table 3.1: Generated items for Clear and Unambiguous

Clear	Unambiguous
This HMI presents information in a clear way	The information presented could mean multiple things to me
The HMI presents information in a way that is easy to understand	The information presented was distinct
I found it difficult to interpret the information	I can interpret this information in only one way
I did not understand what the message/information meant	It is difficult to confuse this information with something else
The information was not presented in a legible way	I was confused by this information
The information presented was confusing	I understood what was meant by this information
The HMI presents information in a simple manner	I had no doubt about the meaning of this information
The HMI presents information concisely	I was uncertain about what this information meant
The HMI presents information in way that is easy to process	The information presented by this HMI was vague
The Information presented by this HMI is easy to perceive	It is easy to misunderstand this information
The HMI presents information coherently	I quickly comprehended this information
The HMI presents information distinctly	This information conflicts with other information

Table 3.2: List of 15 items remaining after expert sessions

Item number	Item text
Q1	I understand what is meant by this information.
Q2	The information is easy to understand
Q3	The information is unclear
Q4	I can quickly comprehend this information
Q5	I can interpret this information in multiple ways
Q6	The information is presented clearly
Q7	I find it easy to interpret the information
Q8	I am uncertain about what this information may mean
Q9	The information is presented in a clear manner
Q10	The information is easy to perceive
Q11	The information is confusing
Q12	The information is legible
Q13	I am confused by this information
Q14	The information is ambiguous
Q15	The information can have multiple meanings

4 Survey Study

4.1 Introduction

The Survey study was intended to further develop a rating scale that can measure the construct of “Clear and Unambiguous” for automotive HMIs. The initial set of items generated in section 3.3 was therefore presented to a representative sample of respondents in order to evaluate reliability and internal consistency, identify potential underlying factors and to reduce the set of 15 items to a smaller set of reliable items.

4.2 Method

4.2.1 Participants

Members of a Dutch TNO participant database were approached via e-mail to partake in an online questionnaire built using Survalyzer. The first 100 participants received 15 Euros in compensation.

Participants were included based on the following criteria:

- Between 18-65 years of age
- In possession of a valid Dutch driver’s license
- (Corrected to) normal eye sight (colour-blind participants were excluded from the study).

Participant inclusion was based on their self-report concerning these criteria.

A total of 104 complete responses were received. Five responses showed duplicate entries (e.g., identical IP-addresses with identical or highly similar responses on the second entry). The second entries of these five participants were removed from further data analyses. The data from the remaining 99 participants (mean age = 38.9, SD = 13.9; 57 females) were used in subsequent analyses after checking for potential outliers.

4.2.2 Stimuli

Six stimuli were created for application of the questionnaire. The stimuli consisted of images of a dashboard interface, using the dashboard of a Honda Accord as a basis², from which the green dashboard lighting and icons not needed for the survey were omitted. This dashboard was modified to show messages on Cruise Control, Lane Centering, Adaptive Cruise Control, Forward Collision Warning, Lane Departure Warning, or Lane Keep Assist. These messages were intentionally created with varying levels of quality, with the goal of showing a wide range of stimuli. The HMIs for Cruise Control, Lane Centering and Adaptive Cruise Control were considered to be quite clear and to produce higher scores on the questionnaire, while those for Forward Collision Warning, Lane Departure Warning and Lane Keep Assist were expected to score lower.

² Downloaded from <https://www.verneidehondasiouxcity.com/what-is-honda-adaptive-cruise-control-acc/>.

For the cruise control stimulus, existing iconography for Cruise Control² was used for familiarity (Figure 4-1). Next to that, the set speed was added in matching green and a green indicator line was added for further clarification.



Figure 4-1. Cruise Control.

The icon for Lane Departure warning³ was used with a contrasting green text above to create confusion (Figure 4-2). The text mentions to 'Stay in lane' and has a white square for drawing attention.



Figure 4-2. Lane Departure Warning.

The icon for Forward Collision Warning⁴ was placed centred in the dashboard (Figure 4-3). Due to the figure having no red or amber colours indicating urgency, this figure could be slightly unclear.

² <https://www.nrautorepair.ca/symbols>

³ <https://mocktheorytest.com/resources/what-does-lane-keep-assist-do/>

⁴ https://www.youtube.com/watch?v=_sKzIgMsE_E



Figure 4-3. Forward Collision Warning.

The Lane Keep Assist icon⁵ was used intentionally without any further explanation and in a small size to create a less clear message (Figure 4-4).



Figure 4-4. Lane Keep Assist.

The figure for Lane Centering was adapted based on existing graphics⁶ with adapted colouring (Figure 4-5). The text above the icon was added to clarify the meaning of the message.

⁵ <https://www.volkswagen.co.uk/en/owners-and-services/my-car/warning-light/electric-lane-keeping-system-lane-assist-not-available.html>

⁶ <https://creazilla.com/nodes/36470-hands-are-driving-the-steering-wheel-clipart>



Figure 4-5. Lane Centering.

The Adaptive Cruise Control icon was used from the base Honda image. Though this image does not contain the ISO logo for Adaptive Cruise Control, the meaning of the functionality could still be partially derived from the icon showing a lead vehicle and distance lines (Figure 4-6).



Figure 4-6. Adaptive Cruise Control.

4.2.3 Procedure

Upon opening the online questionnaire and signing the informed consent form, participants were presented with a brief introduction. This introduction described how in the near future cars will likely take over more driving tasks from the human. As such, it is important that the car's HMI conveys which tasks are performed by the automation. Participants were further informed that they were to answer questions on a 7-point Likert scale (1: *Strongly disagree*, to 7: *Strongly agree*) for six instrument clusters showing one of the following six Advanced Driver Assistance Systems being active:

- Adaptive Cruise Control (ACC)
- Lane Centering (LC)
- Cruise Control (CC)
- Forward Collision Warning (FCW)

- Lane Departure Warning (LDW)
- Lane Keeping Assist (LKA)

After the introduction and after providing background information to confirm inclusion criteria, participants were presented with a page showing:

- A brief verbal description of one of the six ADASs
- An image of a dashboard with the respective ADAS' notification
- The 15 items generated in section 3.3.

All 6 ADASs and items were presented in a randomized order.

Lastly, participants were asked to rate their familiarity with the six ADASs.

4.3 Results

4.3.1 Manipulation check

To determine if the six designed dashboards were indeed perceived as different, a repeated measures ANOVA was performed by comparing the effect of HMI on the mean ratings of all 15 items (see [Table 4.1](#)).

Table 4.1: Mean ratings per HMI ('Std. Error' refers to standard error of the mean).

Repeated Measures ANOVA				
HMI	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	4.87	0.13	4.61	5.13
2	5.21	0.14	4.94	5.48
3	5.73	0.10	5.53	5.94
4	5.51	0.098	5.32	5.70
5	5.37	0.11	5.15	5.60
6	4.40	0.13	4.15	4.66

There was a statistically significant difference in mean ratings ($F(4.78, 468.81) = 21.54, p < 0.001$). Pairwise (Bonferroni corrected) comparisons revealed that each HMI significantly differed in ratings with at least one other HMI (see [Table 4.2](#)). As such, we considered the manipulations as successful and included all stimuli in further analyses.

Table 4.2: Pairwise comparison between all HMI's on mean rating. Significant differences marked with an *. 'Std. Error' refers to standard error of the mean.

Pairwise Comparisons						
HMI	HMI	Mean Difference	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	-0.34	0.15	0.39	-0.79	0.11
	3	-0.86*	0.14	<0.01	-1.27	-0.46
	4	-0.64*	0.14	<0.01	-1.05	-0.22
	5	-0.50*	0.15	0.01	-0.94	-0.06
	6	0.47	0.17	0.09	-0.03	0.97
2	1	0.34	0.15	0.39	-0.11	0.79
	3	-0.52*	0.15	0.01	-0.97	-0.07
	4	-0.30	0.16	1.00	-0.78	0.19
	5	-0.16	0.15	1.00	-0.60	0.28
	6	0.81*	0.17	<0.01	0.28	1.33
3	1	0.86*	0.14	<0.01	0.46	1.27
	2	0.52*	0.15	0.010	0.07	0.97
	4	0.22	0.12	1.00	-0.15	0.60
	5	0.36	0.14	0.18	-0.06	0.78
	6	1.33*	0.14	<0.01	0.90	1.76
4	1	0.64*	0.14	<0.01	0.22	1.05
	2	0.30	0.16	1.00	-0.19	0.78
	3	-0.22	0.12	1.00	-0.60	0.15
	5	0.14	0.13	1.00	-0.24	0.51
	6	1.10*	0.14	<0.01	0.69	1.52
5	1	0.50*	0.15	0.01	0.06	0.94
	2	0.16	0.15	1.00	-0.28	0.60
	3	-0.36	0.14	0.18	-0.78	0.06
	4	-0.14	0.13	1.00	-0.51	0.24
	6	0.97*	0.15	<0.01	0.53	1.41
6	1	-0.47	0.17	0.09	-0.97	0.03
	2	-0.81*	0.17	<0.01	-1.33	-0.28
	3	-1.33*	0.14	<0.01	-1.76	-0.90
	4	-1.10*	0.14	<0.01	-1.52	-0.69
	5	-0.97*	0.15	<0.01	-1.41	-0.53

4.3.2 Exploratory Factor Analysis

All stimuli were analysed together as this is acceptable for exploratory factor analysis (Bradlow, 2002). The correlation matrix (see [Table 4.3](#)) between all 15 items revealed that several items highly correlated with each other ($r > 0.8$). These items were inspected for similarity in meaning, and the following items were dropped from further analyses to avoid extreme multicollinearity.

- I understand what is meant by this information. [Q1]
- The information is easy to understand [Q2]
- I can interpret this information in multiple ways [Q5]
- I find it easy to interpret the information [Q7]
- The information is presented in a clear manner [Q9]
- I am confused by this information [Q13]

In deciding which items to remove in the case of high correlations, items with simpler wording were preferred over the more complex formulations that were formulated in a less complex way. These decisions were made in team discussions of the results. In particular, the following reasons for item removal were formulated:

- Q1 (“I understand what is meant by this information”), Q2 (“The information is easy to understand”), and Q7 (“I find it easy to interpret the information”) were removed as they highly correlated with Q4 (“I can quickly comprehend this information”). Furthermore, Q4 semantically implied Q1, Q2 and Q7, as a fast understanding of information meant that the information was understood (i.e., Q1), and that it was done so rapidly and therefore easily (i.e., Q2 and Q7).
- Similarly, Q5 (“I can interpret this information in multiple ways”) was removed because it highly correlated with Q15 (“The information can have multiple meanings”) and it was semantically similar. It was decided that Q15 was likely easier to comprehend.
- Q9 (“The information is presented in a clear manner”) was removed as it highly correlated with Q6 (“The information is presented clearly”) and because the latter implied the former.
- Q13 (“I am confused by this information”) was removed as it highly correlated with both Q8 (“I am uncertain about what this information may mean”) and Q11 (“The information is confusing”). Furthermore, Q8 and Q11 together imply Q13.

Table 4.3: Correlation matrix for all items. Highly correlated questions highlighted in orange.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
Q1	-														
Q2	.81	-													
Q3	.71	.72	-												
Q4	.81	.83	.72	-											
Q5	.61	.62	.64	.61	-										
Q6	.66	.72	.65	.77	.50	-									
Q7	.81	.83	.73	.84	.64	.75	-								
Q8	.79	.73	.73	.73	.71	.63	.78	-							
Q9	.66	.74	.67	.77	.49	.83	.75	.60	-						
Q10	.47	.55	.48	.62	.34	.72	.57	.45	.71	-					
Q11	.77	.76	.75	.75	.72	.66	.78	.79	.67	.49	-				
Q12	.42	.48	.43	.52	.28	.66	.51	.38	.66	.72	.42	-			
Q13	.76	.73	.74	.74	.73	.62	.79	.80	.61	.45	.85	.40	-		
Q14	.58	.57	.60	.56	.72	.46	.60	.67	.47	.32	.71	.26	.72	-	
Q15	.66	.65	.64	.65	.81	.55	.68	.72	.54	.39	.74	.33	.71	.72	-

Removal of these six questions resulted in the following set of items, with inter-item correlations below 0.8.

- The information is unclear [Q3]
- I can quickly comprehend this information [Q4]
- The information is presented clearly [Q6]
- I am uncertain about what this information may mean [Q8]
- The information is easy to perceive [Q10]
- The information is confusing [Q11]
- The information is legible [Q12]
- The information is ambiguous [Q14]
- The information can have multiple meanings [Q15]

Maximum likelihood factoring was conducted on these nine items with oblique rotation (direct oblimin) to allow for correlated factors and for generalizing results to a larger population. The Kaiser-Meyer-Olkin measure verified that the data sample was adequate for analysis (KMO = .92).

An initial analysis was run keeping factors with eigenvalues above 1. Two factors were extracted that together explained 72% of the variance. The scree plot and a low eigenvalue of the 3rd factor (0.45) justified the choice to not include more factors. Table 4.4 shows the factor loadings presented by the pattern matrix (loadings < .35 are suppressed). The items that cluster on the same factor suggest that the first factor (all loadings > 0.59) represents a cognitive dimension of information processing, whereas the second factor (all loadings > 0.67) represents a perceptual dimension.

Table 4.4: Pattern Matrix with factor loadings when extracting two factors ranked highest to lowest

Item	Factor 1	Factor 2
The information is ambiguous [Q14]	.86	
The information is confusing [Q11]	.86	
I am uncertain about what this information may mean [Q8]	.85	
The information can have multiple meanings [Q15]	.85	
The information is unclear [Q3]	.72	
I can quickly comprehend this information [Q4]	.59	.39
The information is easy to perceive [Q10]		.86
The information is legible [Q12]		.84
The information is presented clearly [Q6]		.67

Notably, the item “I can quickly comprehend this information” [Q4] partially loads on the second factor as well. This is not surprising as speed of processing can also be attributed to the ability to quickly recognize (visual) elements and group them together into a meaningful whole (Johnston, 1985; Winkielman, 2003; Topolinski, 2009).

Although two factors were extracted from the initial analysis, clear and unambiguous information presentation has cognitive and perceptual components. As such, it can be argued that the two factors can be taken together as one measure for clear and unambiguous. A second analysis was performed with the constraint to extract only one factor. The single factor explained 60% of the variance with all loadings >.55 (see [Table 4.5](#)).

Table 4.5: Pattern Matrix with factor loadings when extracting one factor ranked highest to lowest

Item	Factor 1
The information is confusing [Q11]	.89
I can quickly comprehend this information [Q4]	.86
I am uncertain about what this information may mean [Q8]	.86
The information is unclear [Q3]	.83
The information is presented clearly [Q6]	.79
The information can have multiple meanings [Q15]	.79
The information is ambiguous [Q14]	.73
The information is easy to perceive [Q10]	.62
The information is legible [Q12]	.55

Notably, the previously double loading item “I can quickly comprehend this information” [Q4] became the 2nd strongest loading item for this factor, warranting the necessity to include this item as a possible measure for clear and unambiguous.

Reliability was calculated for both analyses outcomes using Cronbach’s alpha (see [Table 4.6](#)). The double loading factor “I can quickly comprehend this information” [Q4] was included in both the Cognitive and Perceptual factor on theoretical grounds and since Cronbach’s alpha decreased when removing this item from either factors.

Table 4.6: Cronbach's α for each factor indicated very high reliability for each scale.

Factors	Cronbach's α
Combined (n=9)	0.93
Cognitive (n=6)	0.93
Perceptual (n=4)	0.89

4.4 Discussion

The Survey study aimed to develop a rating scale measuring to what degree an HMI is “Clear and Unambiguous”. The results from the study identified a set of 9 items that can measure this construct. Furthermore, we showed that these 9 items load on two factors that together determine how clear and unambiguous an HMI is considered to be. One factor is more cognitive in nature, and refers to the ease with which information is comprehended. Clarity of information, unambiguity, and understanding of the information are aspects that positively contribute to this factor. The second factor is perceptually oriented. The items belonging to this factor focus on ease of perception, legibility and whether information is presented in clear way (e.g., well organized). Many of the conditions described in section 2.4 likely contribute to a positive evaluation on this perceptual factor. Although we describe these two factors as separate, there is an inherent relation between the two. Perceptual features determine whether people can extract meaningful information from a message. As such, measuring if an HMI is clear and unambiguous requires measuring both factors and considering the scores jointly (i.e., averaging all 9 items) and separately (i.e., averaging per factor). This provides the most insight in why an HMI might or might not be clear and unambiguous.

While this study provides a first step in developing a questionnaire to measure Clear and Unambiguous for automotive HMIs, one limitation was the limited variation in ADAS designs, as well as the static display of these HMIs. The next study aims to extend the results by creating systematic variations of HMIs that are presented in a dynamic context. Furthermore, providing the rating scale in the subsequent experiment allows for measuring test-retest reliability.

5 Experimental tests and results

5.1 Introduction

Our online survey (Chapter 4) showed that our new Clear and Unambiguous Questionnaire reliably measured to which extent static HMIs for various ADAS were considered clear and unambiguous by participants. In practice, however, information also needs to be clear and unambiguous in the dynamic driving context. While it was not feasible within the scope of this project to test the questionnaire during actual driving, we conducted an experimental study in which participants were presented with movie clips of synthetic automation transitions (from manual or assisted driving to automated driving or vice versa). The Clear and Unambiguous Questionnaire was again used to assess to which extent the information provided by the system HMI to the user was rated as clear and unambiguous. In addition, open questions as well as questions regarding the user responsibilities in the presented situations were used to gauge participants' objective understanding of the information. Participants' appreciation of the different HMIs was evaluated using Product Reaction Cards.

5.2 Methods

5.2.1 Participants

Participants were recruited in two separate groups, each consisting of 12 participants. The first group was recruited from students at the Technical University of Eindhoven (TU/e) and was also tested at the TU/e location, while the second group was recruited from the TNO participant database and was tested at the TNO location in Soesterberg (SB). This way, variability in age, driving experience and affinity for technology was expected to be maximized between participants. One of the participants in the second group did not show up for the planned test session. Consequently, in total 23 participants took part in the experiment. They all signed an informed consent before taking part and were paid for their participation. Table 5.1 shows the participants' gender distribution and age. More females than males took part and the TU/e group was, as expected, younger than the Soesterberg group. Participants in the Soesterberg group held their driving license for a longer period of time and also drove more frequently than the TU/e group (which matched the age difference between the two groups).

Table 5.1. Participant characteristics. The Age column shows mean (sd) age in years.

Group	N	F/M	Age
TU/e	12	7/5	24.83 (1.85)
SB	11	8/3	41.82 (16.18)
Total	23	15/8	32.96 (14.00)

5.2.2 Stimuli

Participants were presented with movie clips (duration 1 min) that represented different transitions from manual or assisted driving to automated driving or vice versa (see Figure 5-1). Each movie clip showed a real traffic situation through the windshield in the upper part of the screen and a vehicle steering wheel and dashboard in the lower part (see Appendix D for example screenshots). The movie clips for two transitions (from level 0 to 3 and from level 3 to 0) also included acoustic warning sounds. Each clip started by showing a driving situation on the highway. After some seconds, an icon accompanied by text in the instrument cluster indicated a possible or required transition of control. Additionally, buttons at the 3 o'clock and 9 o'clock positions on the steering wheel blinked. In transition A, the initial situation showed automated driving and the HMI required the user to take over control due to a wet road surface. In transition B, the initial situation showed manual driving and the HMI indicated that automated driving was possible. In transition C, the initial situation showed assisted driving (adaptive cruise control active) and indicated that a transition to automated driving was possible. In all three cases, a text overlay indicated that the transition took place, after which the transition was confirmed by a change in HMI icons.

	Transition	1. Clear	2. Unclear
A	Level 3 → 0	Control Monitoring 	Control Monitoring 
B	Level 0 → 3	Control Monitoring 	Control Monitoring 
C	Level 2 → 3	Control Monitoring 	Control Monitoring 

Figure 5-1. Automation transitions (A-C) and versions (1. Clear; 2. Unclear) used in the experiment. The transitions were simulated to take place between different driving automation levels (SAE, 2021): 0 = manual driving; 2 = assisted driving with ACC and Lane Centering; 3 = automated driving with Automated Lane Keeping System). After transition A, the human driver was required to both control the vehicle and monitor traffic. After transitions B and C, the driver was not required to do either of these, because the automation system took over the role of driver completely.

All three transitions could be presented in a clear version, which was expected to clearly indicate what happened and what was expected of the human driver, or in an unclear version, which was expected to be far less clear to the driver. Table 5.2 lists the principles that were used or violated in order to create the clear and unclear versions, respectively. In short, the clear versions specified the reason for the transition concretely, contained highly visible elements in the instrument cluster with consistent colour use (red for emergency, amber for warning and green for normal or active status), with multimodal congruency (acoustic signals matching visual information) and adhered to ISO symbols where possible. In contrast, the unclear versions violated these principles in several ways. The transition videos were shown full screen on a 24" Philips Brilliance 241B computer monitor (resolution 1920 * 1080), situated at approximately 40 cm from the desk edge (see Figure 5-2).

Table 5.2. Principles used in the transition videos to create clear and unambiguous or unclear/ambiguous information.

Principle	Clear version	Unclear version
System state explanation	Specific	Absent or abstract
Visibility of elements	High	Low
Colour use	Consistent	Inconsistent
Multimodal info	Congruent	Incongruent
ISO symbols use	Yes	No



Figure 5-2. Test setup. Transition videos were shown on the monitor, questions were answered on the laptop.

5.2.3 Measures

All participants were requested to specify their age and gender, as well as how long they had been in possession of a valid driving license (in categories 'less than 1 year', '1 to 3 years', '4 to 10 years', and 'more than 10 years') and how often they drove ('Never', 'Once a month', 'Once a week', 'Several times a week', '(Almost) every day'). They also indicated how much experience they had ('None', 'Some', 'Extensive') with six different ADAS (Cruise Control, Adaptive Cruise Control, Forward Collision Warning, Lane Departure Warning, Lane Keep Assist and Lane Centering) and filled out the Affinity for Technology Interaction (ATI) scale (Franke, Attig, & Wessel, 2019) in the Dutch version (translation by Nick van Apeldoorn

in the Dignity project). Through a copying mistake, only eight of the nine ATI questions were used (item 7 was not included).

After each transition video, participants answered several questions. Two open questions asked them to describe what they had seen and what they thought the meaning of the information was:

1. Can you describe what happened inside the car?
2. What do you think this means?

The experiment leader wrote down their answers to these questions and asked clarifying questions when needed. These answers served to check whether participants had accurately perceived and understood the HMI information.

Participants also answered the following three questions on a 5-point scale ('Certainly not', 'Probably not', 'Don't know', 'Probably', 'Certainly'):

1. Do you still need to steer (when necessary)?
2. Do you still need to brake (when necessary)?
3. Do you still need to pay attention?

Participants were instructed to answer these questions for the situation after the transition. These answers gave an indication whether they understood the meaning of the information they had received from the HMI in the movie clip.

Participants filled out the Clear and Unambiguous Questionnaire for the HMI they had just seen, answering all 9 questions on a 7-point scale (see Table 5.3). This measured their subjective rating of the HMI information.

Table 5.3. Clear and Unambiguous Questionnaire

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 final measure was the choice of a set of three out of twenty Product Reaction Cards (Benedek & Miner, 2002). The Product Reaction Cards were originally developed by Microsoft

and consist of 118 terms that might be used to evaluate a product. The 20 terms thought to be most appropriate for our HMI stimuli were selected from the complete set and translated into Dutch. Figure 5-3 shows the 20 words in English. The Dutch words as shown on paper to the participants are given in Appendix B. Participants were asked to pick the three terms that they thought most fitting for the HMI they had just seen in the movie clip. The experiment leader entered these into the laptop. All questions were posed in Dutch.

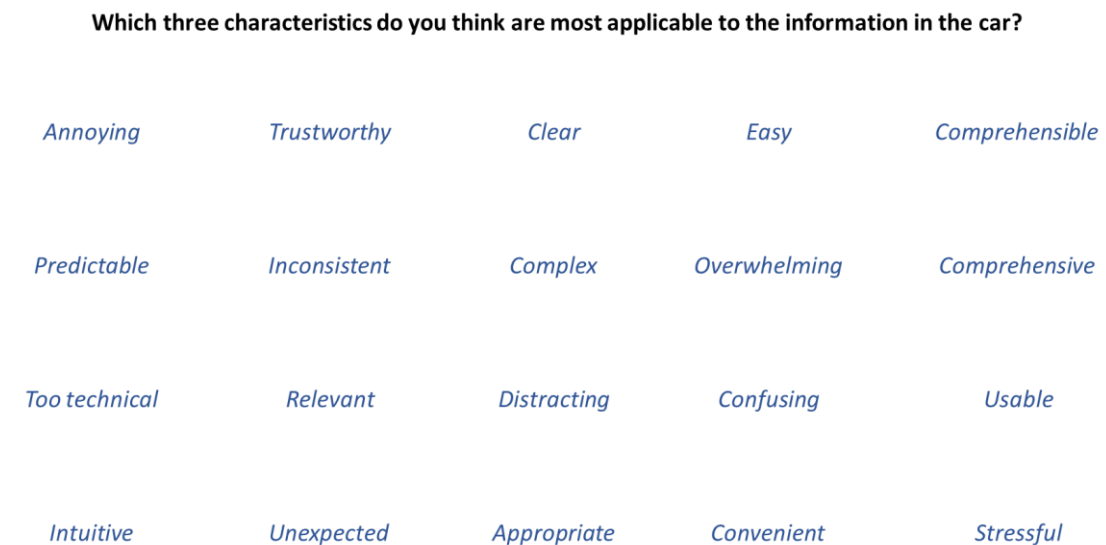


Figure 5-3. English version of the Product Reaction Cards used in the experiment.

5.2.4 Design and procedure

Each participant saw one version (clear or unclear) of each transition (A, B, and C). Thus, a mixed design in which transition was varied within participant but version between participants was used. This way, the variation caused by different transitions could be analysed, while preventing exposure to one version of a given transition influencing responses to the other version. The order of transitions was counterbalanced across participants within each participant group (except for the one missing participant in the SB group). Half the participants received one clear and two unclear HMIs, while the other half received two clear and one unclear version. Care was taken that in both participant groups, half the transitions were shown in the clear and half in the unclear version across all participants in that group.

Each test session started with an introduction of the experiment and signing of the informed consent by both participant and experiment leader (see Figure 5-4). The introduction included a brief description of new (e.g., ALKS) or future (e.g., ALKS+) driving automation systems, which allow the user to perform non-driving related activities, without paying attention to the driving task. The participant was informed that he/she would see movie clips with transitions from or to automated driving. The participant also answered the questions concerning age, gender, driving license and driving experience. After this, the participant was shown an example video to explain what the videos they were going to see for the three transitions looked like. When participants had no further questions, they were shown the three transition videos and each time asked the same questions (see above, section 5.2.3). Each video was shown twice in a row, to reduce the memory burden put on

participants. Each block of transition video and questions took approximately 8 minutes. After the last block, participants indicated their experience with ADAS and filled out the ATI to end the session. The total procedure took about 30 – 40 min.

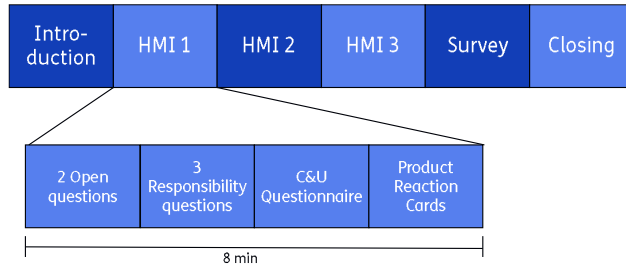


Figure 5-4. Experimental procedure.

5.2.5 Analysis

5.2.5.1 Open questions

In order to score the participants' descriptions of the HMIs, first for each transition movie the HMI elements that were present were classified according to four categories:

1. Icons in the instrument cluster
2. Text in the instrument cluster
3. Buttons on the steering wheel
4. Sounds

For each of these categories, the elements present in each transition movie were determined. Two independent raters then scored the answers from all participants by counting the number of elements correctly described by the participants. Only correct elements were counted, no points were subtracted for incorrectly mentioned elements. The scores from the two raters were averaged to determine a participant's score and divided by the maximum possible score, producing a score between 0 and 1.

For the answers concerning the meaning of the HMI information a similar procedure was followed. For each transition movie, first the correct start state, the end state or transition and the reason given for the transition were defined. Then, two independent raters scored the participants' answers by counting how many of these elements were correctly mentioned by each participant. The scores from the two raters were averaged to determine a participant's score and divided by the maximum possible score, producing a score between 0 and 1.

5.2.5.2 Responsibility questions

The correct answers to all three questions (concerning steering, braking and paying attention) were 'Certainly' for transition A (from automated to manual driving) and 'Certainly not' for the other two transitions towards automated driving. Answers were scored -2 for 'Certainly not', -1 for 'Probably not', 0 for 'Don't know', 1 for 'Probably' and 2 for 'Certainly'. For transitions B and C, scores were inverted such that the correct answer 'Certainly not' led to the maximum score of 2. For final analysis, the average of the scores on the three questions was taken.

5.2.5.3 Clear and Unambiguous Questionnaire

Answers on the 7-point scale were scored from -3 to +3, with the central answer category ('Agree nor disagree') corresponding to 0. Scores on negatively phrased items were inverted before further analysis. Answer distributions for individual items were inspected visually for outliers or deviant patterns, using the 'likert' package⁷ in R (version 4.2.1). In order to quantify reliability and internal consistency, Cronbach's Alpha and Guttman's lambda_6 were computed using R with the 'psych' package⁸. In order to evaluate the difference in questionnaire scores between the clear and unclear versions of the three transitions, first two factor scores for the questionnaire were computed by weighing scores on individual items by each item's factor loadings (see [Table 4.4](#)) and computing weighted averages for the perception and comprehension factors. These were then entered into a linear mixed-effects model (LMM) analysis performed with the R package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). Transition (A, B, or C), version (clear or unclear) and their interaction were included as fixed effects, ATI score was included as a continuous factor and the model also included random intercepts for participants. The Restricted Maximum Likelihood estimation method was used.

5.2.5.4 Product Reaction Cards

The frequency of the terms chosen by the participants for each transition was counted and represented graphically. The chosen terms were also categorized into positive and negative and the number of positive/negative terms was compared for the three transitions in both versions (clear / unclear).

5.3 Results

5.3.1 Experience with ADAS and affinity for technology

Figure 5-5 shows the average participant scores on ADAS experience for both participant groups. In general, participants had limited experience with ADAS (mainly with traditional cruise control). This was especially true for the Soesterberg group. ATI scores correlated strongly with age ($r = -0.62$, $p = 0.0015$). Older participants generally scored lower on the ATI than younger participants (Figure 5-6).

⁷ <https://cran.r-project.org/web/packages/likert/>

⁸ [The Personality Project's Guide to R \(personality-project.org\)](#)

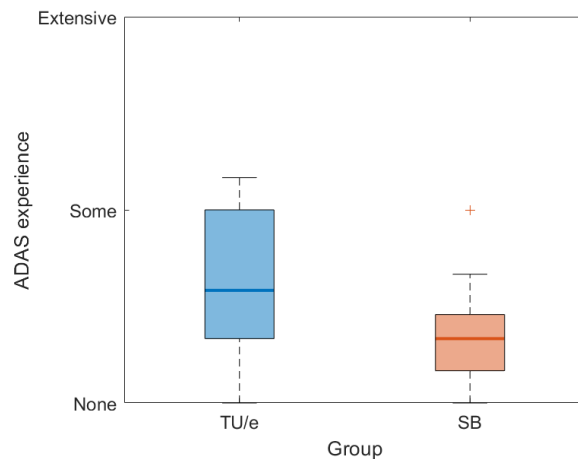


Figure 5-5. ADAS experience in the two participant groups.

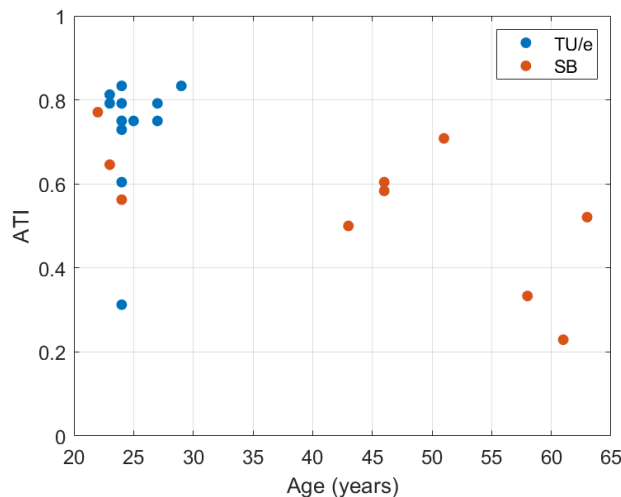


Figure 5-6. Affinity for Technology Interaction score versus participant age for the two participant groups.

5.3.2 Open questions results

Appendix C provides some example answers from participants to the open questions concerning what they had seen and what they thought this meant. In general, participants varied widely in both the quantity and quality of their answers. One thing that stood out from participants' answers was that they often thought the HMI referred to distance keeping. To quantify the agreement between the two raters, Cohen's Kappa was computed. For the scores of the description answers, $K = 0.03$, which indicates that interrater agreement was around chance ($K = 0$). For the scores of the answers about the meaning of the HMI information, $K = 0.53$, indicating better, but not very good agreement. Figure 5-7 shows the results for the answers to both open questions. For participants' descriptions of what they had observed, no clear pattern for different transitions and versions could be seen. For participants' answers concerning the meaning of the information, the clear version of transition A (from L3 to L0) provided more accurate answers, while the unclear version of transition B (from L0 to L3) produced very inaccurate answers. Given the low level of interrater agreement, no further analysis was performed on these scores.

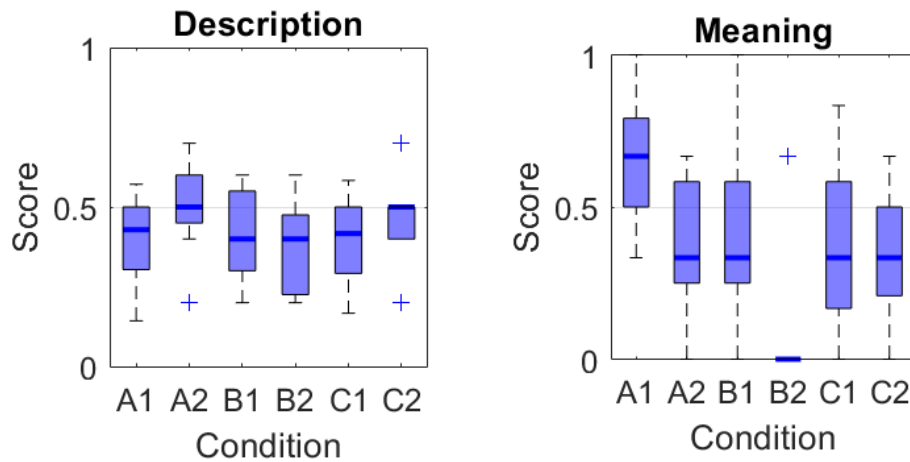


Figure 5-7. Boxplots showing the distributions of participants' answers to the open questions (left: description; right: meaning) as scored by two independent raters. The horizontal axis shows the six conditions defined as combinations of transition (A, B, C) and version (1. Clear; 2. Unclear).

5.3.3 Responsibility questions

Figure 5-8 shows the results for the three questions concerning whether participants thought they still had to steer, brake or pay attention to traffic at the end of each transition, as well as the average score of these three. In general, scores were higher (more correct) for transition A (from L3 to L0), which corresponds with the participants assuming that they always should be ready to steer, brake and pay attention. The scores did not clearly discriminate between the clear and unclear versions of the transitions.

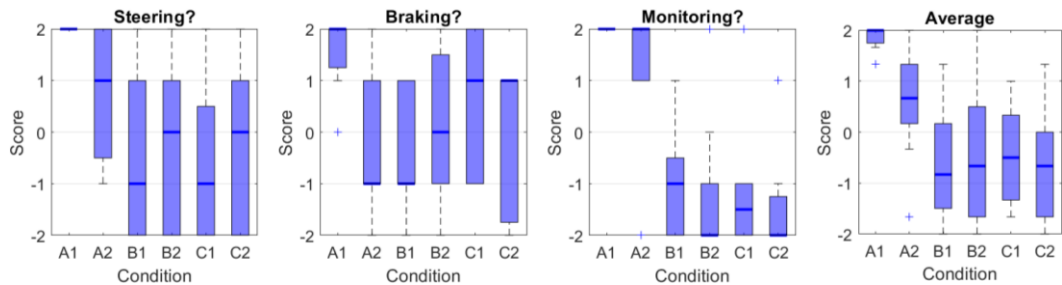


Figure 5-8. Boxplots showing the distributions of participants' answers to the responsibility questions concerning steering, braking and monitoring traffic, as well as the averages of these three. The horizontal axis shows the six conditions defined as combinations of transition (A, B, C) and version (1. Clear; 2. Unclear).

5.3.4 Clear and Unambiguous Questionnaire

Cronbach's alpha for the Clear and Unambiguous Questionnaire was $\alpha = 0.88$ (95% ci: [0.84, 0.92]), with Guttman's $\lambda_6 = 0.91$, indicating high internal consistency. Figure 5-9 shows the participant responses per Clear and Unambiguous Questionnaire item, collapsed across the three transitions. It is evident from the figure that participants systematically rated the clear version of the transitions higher than the unclear version. This was confirmed by the LMM analysis. After aggregating individual item responses into the two factors (perception and comprehension), a significant effect of version on both factors was found (see Table 5.4 and Table 5.5). In addition, a significant interaction effect of transition and version was observed

for the comprehension factor, caused by a smaller difference between the clear and unclear versions for transition C. For the perception factor, the fixed factors explained 22% of the variance, while the total model, including the random intercept, explained 38%. For the comprehension factor, these numbers were 33% and 61%, respectively.

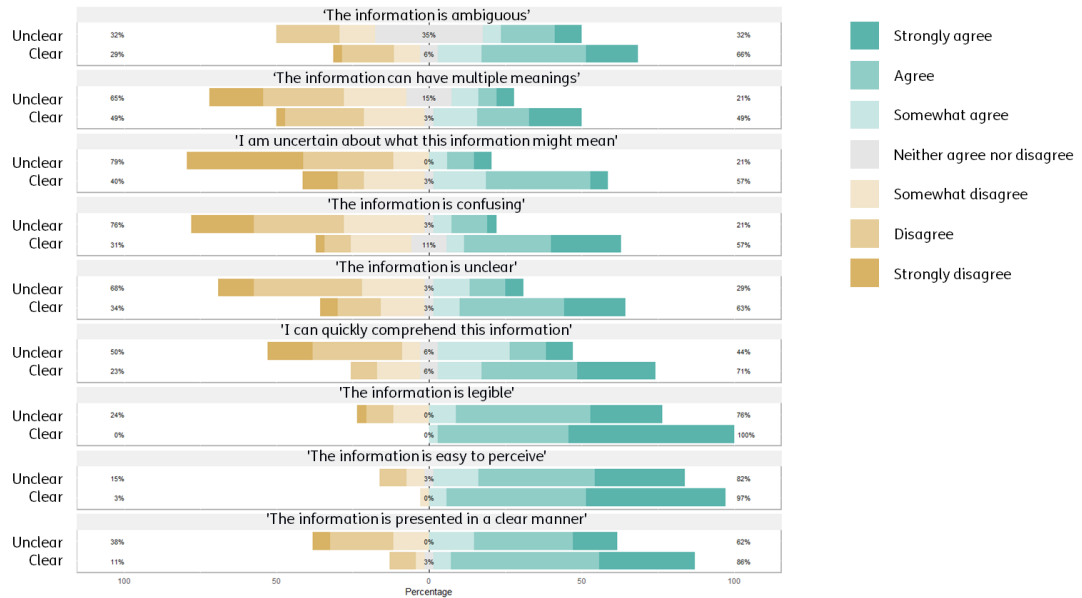


Figure 5-9. Clear and Unambiguous Questionnaire responses. Per item, the two bars show participant responses collapsed across the three different transitions for the clear and unclear versions.

Table 5.4. LMM results for the Perception factor of the Clear and Unambiguous Questionnaire. The notation A:B represents the interaction between factors A and B.

	SS	MS	df1	df2	f	P(F>f)
Transition	5.344	2.672	2	40.493	1.3663	0.2665
Version	34.620	34.620	1	46.010	17.7043	0.0001
ATI	0.062	0.062	1	19.648	0.0318	0.8602
Transition:Version	5.286	2.643	2	61.698	1.3517	0.2664

Table 5.5. LMM results for the Comprehension factor of the Clear and Unambiguous Questionnaire. The notation A:B represents the interaction between factors A and B.

	SS	MS	df1	df2	f	P(F>f)
Transition	0.0373	0.0186	2	40.243	0.0300	0.9705
Version	24.0283	24.0283	1	43.510	38.690	1.671e-07
ATI	0.2462	0.2462	1	19.728	0.3964	0.5362
Transition:Version	6.1000	3.0500	2	57.251	4.9111	0.0108

5.3.5 Product Reaction Cards

Figure 5-10 shows how often the different Product Reaction Card terms were chosen for each transition. In general, more positive terms were chosen for the clear (1) version of the transitions, while more negative terms were chosen for the unclear (2) version. This is confirmed by aggregating the number of positive and negative terms per transition (Figure 5-11).

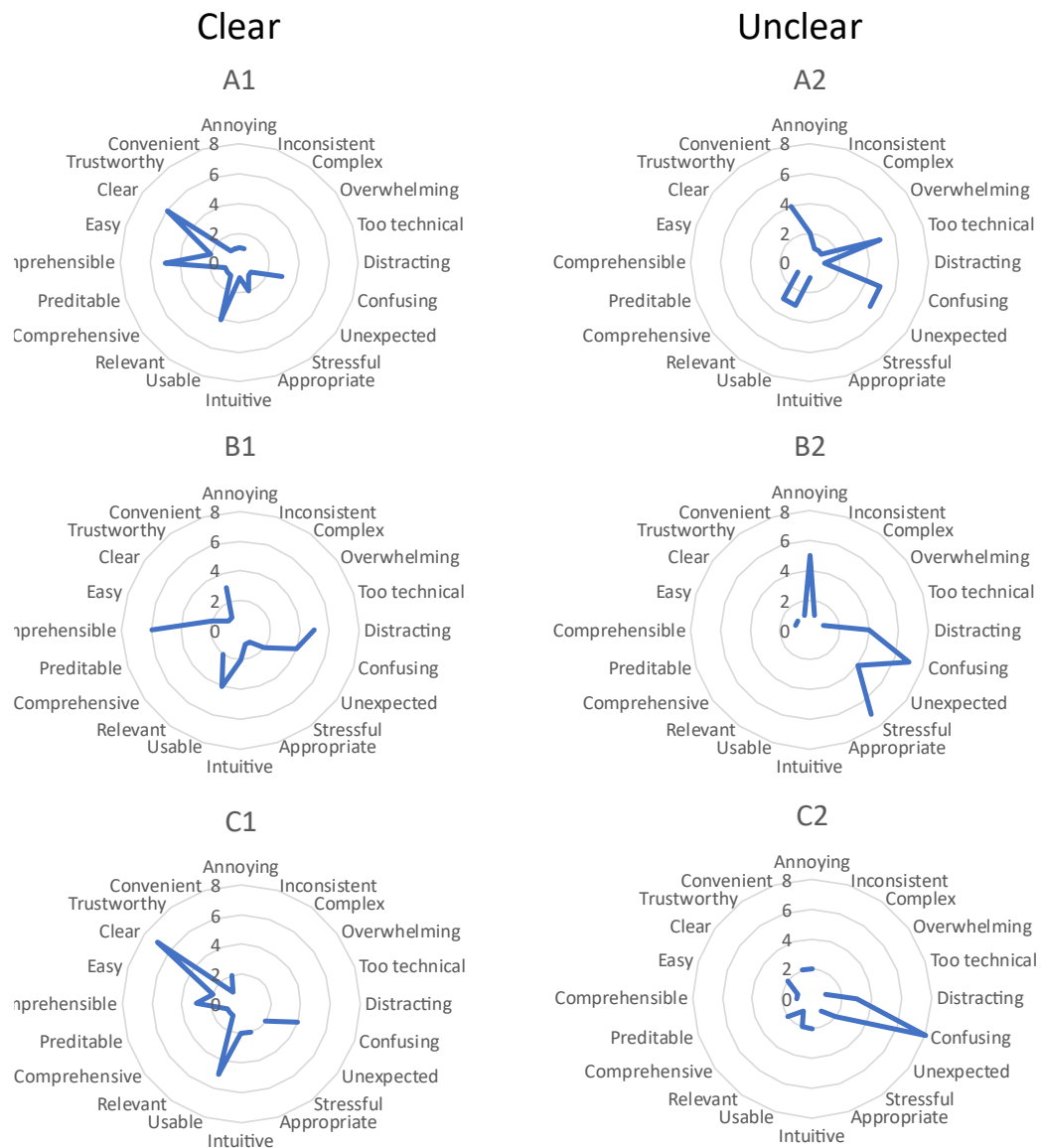


Figure 5-10. Frequency of Product Reaction Card terms chosen for the transitions A, B and C in clear (1) and unclear (2) versions. Each radar plot has positive terms on the left and negative terms on the right.

SUM OF POSITIVE AND NEGATIVE ASSOCIATIONS PER TRANSITION

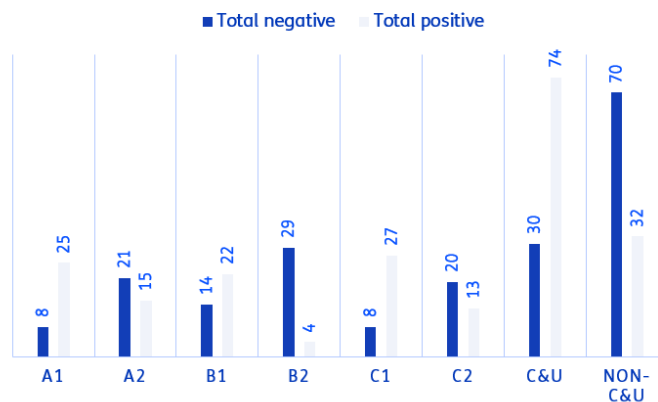


Figure 5-11. Total number of positive and negative terms chosen for each transition. The last 2 columns indicate the number of terms per version (C&U = clear; non-C&U = unclear), which is the sum of the numbers for the individual transitions.

5.4 Discussion

The main goal of the experiment was to investigate whether the new C&U Questionnaire would be able to distinguish between clear and unclear HMIs. Two versions, a clear and an unclear one, were created of three different transition HMIs, simulating transitions between manual or assisted driving and automated driving. Significant differences between the clear and unclear HMIs were indeed observed for the two factors of the questionnaire, perception and comprehension, suggesting that participants rated the information in the clear HMIs as easier to perceive and to understand. However, even for the unclear HMIs participant ratings were on average on the positive side for the comprehension factor, suggesting that the absolute score on the questionnaire factors is not suitable for deciding whether an HMI is clear and unambiguous or not. In part, this may be caused by the lack of a well-defined reference point for when an HMI is to be considered clear and unambiguous. In addition, participants were asked to rate the HMI as a whole. We can, however, not exclude the possibility that they considered elements of the HMI in their answers, rather than the HMI as a whole, or the function it related to. For instance, the impression that several icons were quite clear may have led to a favourable score, even though the overall meaning was unclear. This cannot be established from the current data. In order to understand the relationship between the perception and comprehension of individual components of the HMI and the overall function requires more comprehensive testing, in which the understanding of both components and overall information is measured. In addition, suitable reference points need to be developed, allowing for the development of evaluation criteria based on questionnaire responses. These reference points could be predefined HMIs that are commonly agreed upon to be good/clear HMIs and bad/unclear HMIs. It is currently an open question whether these can be generic reference points, suitable for evaluating any HMI, or that these need to be function specific (e.g., different for an level 3 ALKS HMI than for a level 2 ADAS HMI or a level 4 automation HMI).

It proved far harder to evaluate objectively whether participants had understood the information provided to them by the HMIs. The questions asked (open questions and

responsibility questions) were on purpose designed to be sufficiently general not to depend on the specific HMI or function tested. However, for the answers to the open questions, this meant that quantifying to which extent participants understood the information presented to them was not easy. This was reflected in the poor agreement between the two raters who did the scoring, despite developing an a-priori scoring scheme. The responsibility questions also did not provide a clear picture of participants' understanding, possibly because the questions were too general. Some participants remarked that 'of course' they had to keep paying attention, because that is what they are used to in driving. Apparently, providing them with short information concerning the capabilities of ADS was not sufficient to override this default response. Exposure to actual ADS driving (instead of movie clips), for instance in a driving simulator, before performing tests as done here might change their responses to this type of questions, making it a more valid method to assess participant comprehension. Alternatively, questions specifically geared towards the ADS being tested could be used, allowing for much more detailed testing of participants' comprehension. This has the obvious disadvantage that these questions need to be defined separately for every ADS that has to be tested.

The Product Reaction Cards provided a surprisingly clear difference between the clear and unclear versions of the transition HMIs. This suggests that for some purposes, this method may also represent an easy, quick and intuitive way to probe participants' subjective evaluation of how clear and usable an HMI is. More work is needed to define a generally usable set of terms and to validate this as an HMI evaluation method, rather than a qualitative user experience tool. In particular, similar to the C&U questionnaire, reference points need to be established in order to develop some kind of criterion for HMI evaluation.

Even though we did our best to provide some traffic context in the movie clips and focus on automation mode transitions, still our test was a relatively abstract representation of real driving. In addition, participants received some information about ADS capabilities, but this proved to be insufficient for them to really grasp what this would mean in practice. Thus, while our results show that we can at least subjectively discriminate well between clear and unclear HMIs for ADS, one should be careful in generalizing these results to real driving. More research in real driving situations is needed, in which also the user response to HMI information is considered.

6 Conclusions and recommendations

6.1 Summary

While the introduction of automated driving is intended to make driving safer and more comfortable, it may also make the interaction between user and vehicle more complex and prone to confusion. Users will sometimes be the active driver and at other times function as supervisor of the automation system. These different roles come with different responsibilities for the user in terms of vehicle control and traffic monitoring. This may create unintended safety risks in case the user is not aware of what his/her responsibilities are and what the capabilities of the automation system are. It is therefore important to be able to assess to what extent ADS provide their users with clear and unambiguous information concerning system state and concerning allowed or expected actions from the side of the user. In this project, TNO has developed a questionnaire to measure whether the information provided by an ADS to its user is experienced as clear and unambiguous. The questionnaire consists of nine items, which were shown to discriminate between good and bad HMI designs in terms of user perception and comprehension.

The development and evaluation of this questionnaire was done in four steps. First, a summary was compiled of existing methods and knowledge on the evaluation of HMIs with respect to “clear and unambiguous” and similar concepts. This was done based on a literature review and expert interviews, with a particular focus on automation systems. Knowledge from other relevant domains, such as automation in the medical and military domains as well as in aviation, was also included. The results showed that, while design guidelines and principles are often fairly concrete and specific for the low level perceptual characteristics of HMIs (e.g., use of colour and symbol size), they quickly become more general and abstract when it comes to how well the user can comprehend the information provided by a system. Moreover, quantitative measures of HMI quality are rare and no clear criteria for HMI evaluation are available. Hence, the current state of affairs is that each individual HMI has to be assessed on an ad-hoc basis with methods that need to be defined and implemented specifically for that HMI and are often qualitative rather than quantitative.

The second step within the project was to start defining questionnaire items for a more general HMI evaluation method, which is not only applicable to HMIs for different ADS, but also for a variety of user groups and use contexts. Initial items were formulated based on the literature review and expert interviews and were then ranked by a group of HMI design experts. This resulted in a set of 15 questionnaire items that could be presented as Likert scale items on a 7 point scale.

In order to test the reliability and internal consistency of the questionnaire as well as explore underlying factors, an online survey was conducted as a third step. In total 99 participants evaluated six HMI displays for different ADAS using the set of 15 questionnaire items defined in step 2. The HMIs were designed to vary in the degree to which they provided clear and unambiguous information. Data analysis showed that responses on some of the items were highly correlated. A subset of 9 items was found to be able to discriminate well between

HMIs with different levels of clarity and (un)ambiguity. Exploratory factor analysis revealed two underlying factors, which could be interpreted to relate to perception and comprehension of HMI information.

In the fourth and final step, the 9-item questionnaire was tested in an experimental setup. 23 Participants were presented with movie clips of transitions from or to automated driving, including a view of the traffic situation (through the windshield) and of the instrument cluster and steering wheel. Again, HMIs were designed to be either clear and unambiguous or unclear/ambiguous. The results showed that the questionnaire discriminated reliably between both versions of the HMI. As in the online survey, high reliability and internal consistency were observed. Additional questions intended to measure user comprehension more objectively turned out to be less useful. In contrast, qualitative measurement of user experience by means of Product Reaction Cards also showed clear differences between the two different HMI versions.

6.2 Discussion

The questionnaire developed in this research can be used to measure to what degree an HMI is considered to be clear and unambiguous. That said, there are several factors that should be considered when applying the questionnaire to maximize its validity to real-life situations. We discuss the application, limitations and implementation of the questionnaire in the wider design and its potential role in the type approval process below.

6.2.1 User (groups)

Users evaluating an HMI should be representative of the potential users of the HMI. This implies also considering users that might be outside of the targeted audience of an OEM. Demographics, physical (dis)abilities, cognitive (dis)abilities, experience in driving with and without assistance and automation systems, and other known human factors influencing perception and cognition should be considered when determining the user sample. Furthermore, we recommend splitting users in different groups for at least some of these factors to confirm that an HMI under test is evaluated accurately for users with different characteristics. For example, users with no ADAS experience might score an HMI with high automation level features as less clear or more ambiguous compared to users with ADAS experience. Without splitting these users into two groups, there is a risk that such a finding might be averaged out, thereby creating the illusion that a novel HMI is clear and unambiguous for a broad audience.

6.2.2 Towards task performance

As discussed in Section 3.1.2, perception and comprehension of information are only parts of the factors leading to a safe response by a user. A user might still incorrectly respond to a situation even after correctly interpreting information presented by a clear and unambiguous HMI. Conversely, a user might correctly respond to a situation despite an HMI presenting information in an unclear or ambiguous manner. Additionally, driving safety also heavily depends on the user's response time to a potentially critical situation. Although there are general guidelines on timely responses or allowed distraction times (NHTSA, 2016), different driving tasks and situations might warrant different time criteria. The developed questionnaire does not take into account a user's response to the presented information or its latency. Nonetheless, information presented in a clear and unambiguous manner supports users in acting correctly and timely (or correctly refraining from acting) to

a situation. Sections 2.2 and 2.4 specifically summarize design considerations that should positively contribute to the creation of clear and unambiguous HMIs. 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. The tasks used during such an HMI evaluation should be representative of real-life. Furthermore, there should be a specific focus on tasks that were identified as being high-risk (i.e., there is a relatively high likelihood that users might misinterpret certain information displayed on an HMI) and/or can lead to serious harm (including harm to the user, passengers or other road users). It is expected that tasks involving mode switching, such as transitioning from or to automated driving, are risk sensitive and should be included to validate if an HMI is clear and unambiguous. The exact means to identify these risks and tasks are out of scope of this research, but we suggest following a use-related risk analysis approach.

The interaction with an HMI usually takes place in a dynamic traffic environment. HMIs should not only be tested in driving scenarios with low complexity, but also scenarios that demand more attentional resources from the driver, putting more strict requirements on the information provided by the HMI. Moreover, environmental factors can be diverse and influence the user's ability to interact with an HMI. Examples of such factors are acoustic noise, conversations with or between passengers, lighting, vibrations, road distractions or other tasks (driving-related or not) warranting attention. Such factors can degrade the ability of a user to perceive and/or comprehend information provided by an HMI. As such, these factors need to be included or simulated when aiming to thoroughly evaluate if an HMI is clear and unambiguous.

6.2.3 Information to users

OEMs might offer multiple ways to promote user understanding of HMI information besides designing a clear and unambiguous HMI. One prominent information channel is that of the user manual, be it physical or digital. Although it cannot be guaranteed that all users read and understand the vehicle manual, this source of information should also be taken into account when evaluating how well users understand an HMI. This could be done by testing user knowledge of the manual or by testing HMI understanding with and without access to the manual.

6.2.4 Relative rating of clear and unambiguous

The nature of a subjective questionnaire implies that ratings of clear and unambiguous can only be interpreted in relation to other HMIs. As such, there is no clear cut-off where HMIs scoring above this cut-off are considered to be sufficiently clear and unambiguous. An attempt to approach such a cut-off can be made by developing a benchmark HMI that is scored by users and to which the new HMI under test is compared. An HMI score below the benchmark HMI score could be considered as insufficiently clear and unambiguous.

6.3 Recommendations

SWOV and TUDelft recently investigated to what extent OEMs currently apply User-Centred Design (UCD) principles in the development of ADAS or ADS (de Goede, Jansen, & van Grondelle, 2023). UCD is defined as a process where the needs and abilities of different types of users are taken into account (see ISO standard 9241-210:2019). The idea is that in the absence of unambiguous system requirements for the interaction between an ADAS/ADS

and its users, evaluating the application of a UCD process could provide an alternative way of auditing safe system interaction. Based on a series of interviews with OEMs, the authors concluded that while the industry considers UCD a valuable approach in ADAS/ADS development, no indications were found that UCD principles were applied all the time. When they *were* applied, they seemed to focus more on customer satisfaction than on human factors in general. The OEMs reacted critically to the idea of a product-based UCD audit. There was moderate support for periodic UCD certification on a company/department level.

Without consensus on the role of UCD-based audits for ADAS/ADS, it becomes all the more important to have agreed upon test criteria. The questionnaire developed in this project can serve as a first step in developing these. At the same time, this questionnaire can be applied throughout the development of ADAS/ADS: to choose between alternatives early in the design process, as well as to validate a final design.

Concerning further development of the C&U Questionnaire, we recommend the following steps:

- Development of a benchmark HMI (or HMIs), which can function as a reference point for questionnaire results on the HMI under evaluation. The average score on the benchmark HMI in a representative user sample could serve as an assessment criterion.
- Development of a test procedure that includes user responses to HMI information in various traffic scenarios, including potentially high risk ones. These scenarios should include at least transitions from manual to automated driving and vice versa. In the case of transitions from automated back to manual driving, these scenarios should include both planned and unplanned transitions, as well as system-initiated and user-initiated transitions. 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 environmental factors that may negatively affect perception and comprehension of HMI information, such as 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.
- The most important dimensions of user diversity need to be identified and minimum requirements for HMI testing defined. This should at least include experienced users familiar with the system under evaluation and new users without experience.

As the C&U questionnaire only measures subjective evaluation of how clear and unambiguous an HMI is, it is important to evaluate the relationship between questionnaire results and driver performance with a given HMI. This links to the much broader question of how to assess driving safety. If questionnaire scores turn out to be highly predictive of driver responses to HMI information, the questionnaire may be sufficient for HMI evaluation. However, if this is not the case, more objective measures of the impact of HMI design on driving performance may be needed. This may imply an impossibility of an assessment tool that is generally applicable and require the development of dedicated tests for each new HMI or ADS.

7 References

- Association for the Advancement of Medical Instrumentation. (2018). *Human factors engineering: design of medical devices*. Washington, DC: American National Standards Institute, Association for the Advancement of Medical Instrumentation.
- Beaujard, F. (2018). Overview of the HEAA method defined by Airbus for Alarm design. *Embedded Real Time Software and Systems*. Toulouse.
- Benedek, J., & Miner, T. (2002). Measuring Desirability: New methods for evaluating desirability in a usability lab setting. *Usability Professionals' Association Conference*.
- Blijlevens, J., Thurgood, C., Hekkert, P., Chen, L., & Leder, H. (2017). The Aesthetic Pleasure in Design Scale: The development of a scale to measure aesthetic pleasure for designed artifacts. *Psychology of Aesthetics, Creativity, and the Arts*, 86.
- Bradlow, W. (2002). Exploring repeated measures data sets for key features using Principal Component Analysis. *International Journal of Research in Marketing*, 167-179.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 4-7.
- Campbell, J. L., Brown, J. L., Graving, J. S., Richard, C. M., Lichty, M. G., Bacon, L. P., . . . Sanquist, T. (2018). *Human Factors Design Guidance for Level 2 and Level 3 Automated Driving Concepts*. Washington DC, USA: NHTSA DOT HS 812555.
- Campbell, J. L., Brown, J. L., Graving, J. S., Richard, C. M., Lichty, M. G., Sanquist, T., . . . Morgan, J. F. (2016). *Human Factors Design Guidance For Driver-Vehicle Interfaces*. Washington D.C., USA: NHTSA DOT HS 812 360.
- Campbell, J. L., Hoffmeister, D. H., Kiefer, R. J., Selke, D. J., Green, P., & Richman, J. B. (2004). Comprehension Testing of Active Safety Symbols. *SAE International*, 197-203.
- Carsten, O., & Nilsson, L. (2001). Safety Assessment of Driver Assistance Systems. *European Journal of Transport and Infrastructure Research*, 1(3), 225-243.
- Chin, J., Diehl, V., & Norman, K. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 213-218).
- de Goede, M., Jansen, R., & van Grondelle, E. (2023). *User-centred design for type approval of AD(A)S. Roadmap towards a process audit (R-2022-16)*. The Hague, Netherlands: Stichting Wetenschappelijk Onderzoek Verkeersveiligheid (SWOV).
- Department of Defense. (2020). *MIL-STD-1472H – General human engineering criteria for design and development of military systems, equipment, and facilities*. Military and Government Specs & Standards (Naval Publications and Form Center) (NPFC).
- European Aviation Safety Agency. (2018). *Easy Access Rules for Large Aeroplanes (CS-25) (Amendment 15)*. Retrieved from <https://www.easa.europa.eu/en/document-library/easy-access-rules/easy-access-rules-large-aeroplanes-cs-25>
- European Commission. (2008). *Commission Recommendation of 26 May 2008 on safe and efficient in-vehicle information and communication systems: update of the European Statement of Principles on human-machine interface (2008/653/EC)*. Official Journal of the European Union.
- Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction*, 456-467.
- Johnston, W. D. (1985). Perceptual Fluency and Recognition Judgments. *Journal of Experimental Psychology Learning Memory and Cognition*, 3-11.

-
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26.
- Liu, Y.-C., Figalová, N., & Bengler, K. (2022). Transparency Assessment on Level 2 Automated Vehicle HMIs. *Information* 13 (10), 489.
- Michon, J. A. (1985). A Critical View of Driver Behavior Models: What Do We Know, What Should We Know? In *Human Behavior and Traffic Safety* (pp. 485-524). Boston: Springer.
- Naujoks, F., Hergeth, S., Wiedemann, K., Schömig, N., Forster, Y., & Keinath, A. (2019a). Test procedure for evaluating the human-machine interface of vehicles with automated driving systems. *Traffic Injury Prevention*, 20(sup1), S146-S151.
- Naujoks, F., Wiedemann, K., Schömig, N., Hergeth, S., & Keinath, A. (2019b). Towards guidelines and verification methods for automated vehicle HMIs. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60, 121-136.
- NHTSA. (2016). *Visual-Manual NHTSA Driver Distraction Guidelines for Portable and Aftermarket Devices*. Federal Register.
- Raab, M., Lobinger, B., Hoffmann, S., Pizzera, A., & Laborde, S. (2015). *Performance psychology: Perception, action, cognition, and emotion*. Academic Press.
- SAE. (2016). *J2830_201606 - Process for Comprehension Testing of In-Vehicle Symbols*.
- SAE. (2021). *J3016. The Principles of Operation Framework: A Comprehensive Classification Concept for Automated Driving Functions*.
- Souman, J., van Weperen, M., Hogema, J., Hoedemaeker, M., Westerhuis, F., Stuiver, A., & de Waard, D. (2020). *Human Factors Guidelines for Advanced Driver Assistance Systems and Automated Driving Systems*. The Hague, Netherlands: TNO 2020 R 1216.
- Topolinski, S. S. (2009). The analysis of intuition: Processing fluency and affect in judgements of semantic coherence. *Cognition and Emotion*, 1465-1503.
- United Nations. (2023). Agreement Concerning the Adoption of Harmonized Technical United Nations Regulations for Wheeled Vehicles, Equipment and Parts which can be Fitted and/or be Used on Wheeled Vehicles and the Conditions for Reciprocal Recognition of Approvals Granted on the.
- US Food and Drug Administration. (2016). *Applying Human Factors and Usability Engineering to Medical Devices*. Retrieved from [www.fda.gov](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices): <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices>
- Winkielman, P. S. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. K. Musch, *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189-217).
- Yeh, M., Swider, C., Jin Jo, Y., & Donovan, C. (2016). *Human Factors Considerations in the Design and Evaluation of Flight Deck Displays and Controls*. Washington, DC: Federal Aviation Administration.

8 Abbreviations

ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistance System
ADS	Automated Driving System
ALKS	Automated Lane Keeping System
ATI	Affinity for Technology Interaction
C&U	Clear and unambiguous
CC	Cruise Control
CS	Certification Specification
EASA	European Aviation Safety Agency
EC	European Commission
ESoP	European Statement of Principles
Euro NCAP	European New Car Assessment Programme
FCW	Forward Collision Warning
FDA	Food and Drug Administration
HFE	Human Factors Engineering
HMI	Human-Machine Interface
ISO	International Organization for Standardization
KMO	Kaiser Meyer Olkin
LC	Lane Centering
LDW	Lane Departure Warning
LKA	Lane Keep Assist
LMM	Linear Mixed-Effects Model
ODD	Operational Design Domain
OEM	Original Equipment Manufacturer
SAE	Society of Automotive Engineers
SB	Soesterberg
TNPU	Time Needed for Perceived Understandability
TU Delft	Technical University Delft
TU/e	Eindhoven University of Technology
UCD	User-Centred Design

Appendix A

Dutch Questionnaire Items

Table A 1. Initial set of questionnaire items.

Clear	Unambiguous
This HMI presents information in a clear way	The information presented could mean multiple things to me
The HMI presents information in a way that is easy to understand	The information presented was distinct
I found it difficult to interpret the information	I can interpret this information in only one way
I did not understand what the message/information meant	It is difficult to confuse this information with something else
The information was not presented in a legible way	I was confused by this information
The information presented was confusing	I understood what was meant by this information
The HMI presents information in a simple manner	I had no doubt about the meaning of this information
The HMI presents information concisely	I was uncertain about what this information meant
The HMI presents information in way that is easy to process	The information presented by this HMI was vague
The Information presented by this HMI is easy to perceive	It is easy to misunderstand this information
The HMI presents information coherently	I quickly comprehended this information
The HMI presents information distinctly	This information conflicts with other information

Table A 2. Set of questionnaire items after expert ranking.

Item number	Item text
Q1	Ik begrijp wat er bedoeld wordt met deze informatie
Q2	De informatie is gemakkelijk te begrijpen
Q3	De informatie is onduidelijk
Q4	Ik begrijp deze informatie snel
Q5	Ik kan deze informatie op meerdere manieren interpreteren
Q6	De informatie wordt duidelijk gepresenteerd
Q7	Ik vind het makkelijk om de informatie te interpreteren
Q8	Ik ben onzeker over wat deze informatie kan betekenen
Q9	De informatie wordt op een duidelijke manier gepresenteerd
Q10	De informatie is gemakkelijk waar te nemen
Q11	De informatie is verwarrend
Q12	De informatie is goed leesbaar
Q13	Ik ben verward door deze informatie
Q14	De informatie is dubbelzinnig
Q15	De informatie kan meerdere betekenissen hebben

Table A.1: Refined and Translated list of questionnaire items

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

Appendix B

Product Reaction Cards

The following collection of terms in Dutch was presented to participants on paper:

**Welke drie eigenschappen vind je het meest van
toepassing op de informatie in de auto?**

Vervelend

Betrouwbaar

Duidelijk

Makkelijk

Begrijpelijk

Voorspelbaar

Inconsistent

Complex

Overweldigend

Uitgebreid

Te technisch

Relevant

Afleidend

Verwarrend

Bruikbaar

Intuïtief

Onverwacht

Passend

Handig

Stressvol

Appendix C

Example HMI descriptions

The following tables shows some example answers to the open questions about the transition HMIs that participants were presented with (description and meaning). The first table shows the original, Dutch version, while the second table provides a translation to English.

Condition	Example description	Example interpretation
A1	“Eerst symbool met auto met sensoren eromheen en vervolgens kreeg ik melding dat ik de besturing over moest nemen, zichtbaar op het stuur.”	“De auto kon dus niet meer autonoom rijden en had input van bestuurder nodig om verder te kunnen rijden.”
A2	“Je rijdt, want je ziet een duimpje, dan krijg je auto-error te zien, kennelijk moet je dan het stuur beter vastpakken en de error gaat weg. het tekenetje betekent denk ik dat de auto automatisch afstand houdt van degene voor je, denk ik. maar ik ken het tekenetje niet. er ging een piepje, dat maakte je alert, ik moet wat doen.”	“Waarschijnlijk dat het automatisch systeem iets heeft geregistreerd dat niet helemaal goed gaat en menselijk ingrijpen nodig is.”
B1	“Lampje met bericht licht op waarin stond dat automatisering beschikbaar was, dat kon je activeren door op de knoppen op het stuur te drukken, daarna zag je dat de automatisering actief was.”	“Ik denk dat de auto een functie van mij overneemt, misschien een volg systeem omdat die aan de voor en achterkant die tekenetjes heeft.”
B2	“Een alert van een auto met een pion erop. Dan drukt de bestuurder op de knop.”	“Betekenis is onduidelijk. Geluid wekt idee dat er iets mis is met de auto.”
C1	“Ik zag een kilometer teller met een autootje erboven. Ik zag 2 verschillende afbeeldingen, de eerste zag ik een kilometertje met een autootje en als tweede een autootje voor automatisering. Toen werden de knoppen ingedrukt en toen ging het kilometer tellertje weg en was de automatisering actief.”	“De auto kwam in een situatie dat autonoom rijden mogelijk werd, en bood dit aan aan de bestuurder. De bestuurder ging hiermee akkoord en accepteerde door stuur vast te pakken.”
C2	“Zonder geluid een melding voor automatisering. Stuur licht vervolgens op en wordt ingedrukt.”	“Dat ik hem op automatisch rijden kan inschakelen, dat het daardoor veilig genoeg is.”

Condition	Example description	Example interpretation
A1	"First a symbol with a car with sensors and then I got a message saying that I should take over the steering wheel, displayed on the steering wheel.."	"The car could not drive autonomously anymore and needed driver input to be able to drive on."
A2	"You're driving, because you see a thumb, then you see an 'auto error', apparently you have to grip the steering wheel better and the error disappears. I think the symbol means that the car is maintaining distance to the car in front automatically, but I don't know the symbol. There was a beep, which alerted me that I had to do something."	"Probably the automation system detected something is going wrong and human intervention is needed."
B1	"An icon with tekst message was shown that automation was available, you could activate it by pressing the buttons on the steering wheel, after that you could see that automation was active."	"I think that the car is taking over some function from me, maybe a following system, because the icon has symbols in front and back. "
B2	"An alert of a car with a pylon on top. And then the driver presses the button."	"Meaning is unclear. The sound suggests that something is wrong with the car."
C1	"I saw an odometer with a little car on top. I saw 2 symbols, first a speedometer with a little car and then a small car for automation. Then the buttons were pressed, the speedometer disappeared and automation was active."	"The car entered a situation in which autonomous driving became possible and proposed this tot he driver. The driver agreed and accepted by grabbing the steering wheel.."
C2	"Without sound a message about automation. The steering wheel lights up and is being pressed.."	"That I can activate automated driving, that it is safe to do so."

Appendix D

Experimental test stimuli

The following pages show screenshots from the situation before, during and after the transition of control request for each transition in both the clear (1) and unclear (2) version. In the experiment, these transitions were played as 1 min movie clips. The car interior has been adapted from: <https://www.pexels.com/nl-nl/foto/persoon-met-bmw-stuurwiel-2526128/>.

Stimulus movie A1 (transition from L3 to L0, clear version):



Stimulus movie A2 (transition from L3 to L0, unclear version):



Stimulus movie B1 (transition from L0 to L3, clear version):



Stimulus movie B2 (transition from L0 to L3, unclear version):



Stimulus movie C1 (transition from L2 to L3, clear version):



Stimulus movie C2 (transition from L2 to L3, unclear version):



Mobility & Built Environment

Automotive Campus 30
5708 JZ Helmond
www.tno.nl