CHALMERS

Creating Appropriate Trust for Autonomous Vehicles

A FRAMEWORK FOR HMI DESIGN

Master of Science Thesis in the Master Degree Program Industrial Design Engineering

Fredrick Ekman
Mikael Johansson

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg Sweden, 2015

Department of Product and Production Development
Division of Design & Human Factors

Creating Appropriate Trust for Autonomous Vehicles

A FRAMEWORK FOR HMI DESIGN

FREDRICK EKMAN
MIKAEL JOHANSSON

SUPERVISOR: JANA SOCHOR
EXAMINER: JOHAN HEINERUD

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg Sweden, 2015

Department of Product and Production Development
Division of Design and Human Factors

Master of Science Thesis PPUX05

Creating Appropriate Trust for Autonomous Vehicles
A FRAMEWORK FOR HMI DESIGN
Master of Science Thesis in the Master Degree Program Industrial Design Engineering

Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Telephone: +46 (0)31-772 10 00

Cover photo: Fredrick Ekman
Print: Chalmers Reproservice Göteborg, 2015

ABSTRACT

While autonomous vehicle technology progresses, potentially leading to a safer traffic environment, many challenges remain within the area of human factors. One very important factor that must be addressed is to what extent the driver (user) will be able to trust the self-driving car and its technology. Trust is a cornerstone in the public acceptability of this new and innovative technology.

The aim of the thesis is to explore how an appropriate level of user trust for future autonomous driving vehicles is created. In order to do so, a greater understanding of what affects trust in the Human-Machine Interaction (HMI) in autonomous driving vehicles is studied. This knowledge is then used to generate a guiding framework for implementing trust-related factors into the interaction system. This framework is then used to create an example concept of how a human-machine interaction could be created with regards to trust.

The first main result of the project work is a trust-based framework based on a driving scenario, which can aid future human machine interaction designers in creating interaction systems focused on generating appropriate driver trust. As an initial attempt to corroborate the framework, one feature – object recognition – is tested and validated through a validation test of several test concepts, which confirmed the framework's usefulness in guiding a trust-based development process. The second main result, based on the validation test results, is the further development of an illustrated example concept; demonstrating what types of trust-based interaction system concepts the framework can be used to create.

The authors recommend that HMI designers and autonomous vehicle manufacturers take a more holistic perspective on trust rather than focusing on single, "isolated" events; for example, understanding that trust formation is a dynamic process that starts long before a user's first contact with the system, and continues long thereafter. Furthermore, factors affecting trust change, both during an interaction and over time; thus HMI concepts need to be able to adapt. Future work should be dedicated to understanding how trust-related factors interact with each other, as well as on more comprehensively validating and testing the trust-based framework developed in this thesis.

ACKNOWLEDGEMENTS

Since the area of trust and autonomous vehicles is so extensive a lot of support and feedback was received from experts within the area. These people gave us a lot of inspiration as well as valuable input.

First of all we would like to thank Malin Farmansson, our supervisor at Volvo Car Corporation, for trusting our decisions as well as always supporting us.

Many thanks to Annika Larsson, PhD, and Prof. MariAnne Karlsson for their valuable input regarding the project procedure. Thank you Claes Edgren for the inspiring discussions as well as your ideas.

We also want to thank our examiner Johan Heinerud for always meeting our wishes in best way possible and doing so with a laid-back attitude, calming us down.

Thank you Jonas Olsson, Albin Andersson and Victor Lundgren for your patience and excellent driving during the validation test. We also want to thank all the participants in the user study, validation test and workshop.

And last but not least thank you Jana Sochor, PhD, at Chalmers University of Technology for your excellent supervisory skills and for going the extra mile; without you this project had not been possible within the time frame.

Thank you!

Table of Contents

1 INTRODUCTION

This chapter introduces the thesis project by describing the context, background, aim, goals as well as delimitations made.

The focus of this master's thesis is how to create an appropriate level of user (driver) trust for future autonomous driving (AD) vehicles. Today there are no general guidelines to use when designing an Human-Machine Interaction (HMI) system with focus on trust for autonomous vehicles. In this thesis, creating an appropriate level of trust is done through first producing a guiding framework for implementing trust-related factors into a human-machine interaction system. Then, example HMI concepts are generated, of which one feature is evaluated via a user study. A final, "best" example concept is developed and serves as an illustrative example of how the trust framework can be used.

The structure of the thesis: Following the Introduction (Chapter 1), the Literature Study (2) explores the existing research on trust, autonomous systems and HMI. The Methodology & Implementation (3) chapter then outlines the general design process structure utilized in the thesis. Next, the User Study (4), together with the Literature Study, serves as a basis for developing the trust-based Framework (5) for designing HMI concepts for AD systems. The chapters Concept Development (6) and Validation test (7) then describe the process of generating concepts and testing a feature in order to exemplify how the trust-based framework can be used. The Example Concept (8) chapter uses all parts that could be included into an HMI design for creating trust in order to develop an example concept. The Discussion & Conclusion (9) chapter brings together the thesis' highlights and lessons learned.

This project was initiated in order to better understand how trust can be created through human-machine interaction system design, so that the future driver has an appropriate level of trust for the autonomous driving (AD) vehicles that could be on the roads in the near future. Malin Farmansson, HMI department at Volvo Vehicle Corporation (VCC), and Jana Sochor, Chalmers, Division of Design and Human Factors, acted as the authors' master's thesis supervisors.

1.1 Context

1.1.1 *Technology & Regulations*

As technology within the automotive industry advances, the near future is expected to bring exciting new and innovative systems within the Autonomous Driving (AD) vehicle sector. For example, the vehicle manufacturer BMW shows how their autonomous driving vehicle system could work through a promotional video, letting the vehicle, completely on its own, drift around bends as well as zigzag between obstacles within a closed environment, with the driver as only a passive passenger (BBC, 2014). This type of performance is evidence of how these autonomous systems can perform tasks that earlier only could be performed by humans (Waytz, et al., 2014). Volvo Car Corporation (VCC) is also testing autonomous driving vehicles on the streets of Gothenburg (Volvo Car Group, 2014). Fully autonomous driving is defined as a vehicle that is fully automated and, by that, could drive itself without participation from a driver if the user does not want to have control (van Schijndel-de Nooij, et al., 2010).

Such technical advancements have created an interest from other technology-oriented companies trying to establish themselves in the vehicle market, especially into the AD sector. Since 2010, the high-tech company Google has tested several totally driverless vehicles that have driven on their own for more than 1.6 million km (Pritchard, 2015). Rinspeed in

Switzerland as well as Akka Technologies in France also have their own fully autonomous vehicles, a Tesla-based vehicle called Xchange and a vehicle named Link & Go 2.0, respectively. The Link & Go 2.0 vehicle uses a Global Positioning system (GPS) together with a system called SLAM (Simultaneous Localization Acquisition and Mapping) which is supported by lidar, stereo cameras and other sensors (Frost & Sullivan, 2015).

Enabling systems in AD vehicles range from object detection, radar, laser scanners, communication electronics, artificial vision systems and different software and control algorithms (van Schijndel-de Nooij, et al., 2010). The graphics card manufacturer NVIDIA also tries to incorporate high-end interfaces, updatable HMI system designs as well as sophisticated processor components, creating better Advanced Driver Assistance Systems (ADAS) solutions for future autonomous driving vehicle systems (NVIDIA, 2015). Such systems are fairly new but could lead to both better traffic safety as well as better fuel efficiency in the future (Verberne, et al., 2012).

Although fully and totally self-driving autonomous systems that are updatable and could take you from point A to point B all on their own are not yet implemented in day to day public traffic, other partially autonomous vehicles systems have been implemented and are on the road. These systems allow or rather demand the user to be in control, but the users can still get help from the systems through automation within a lower complexity milieu, for instance on a part of a bigger road such as a motorway (Gasser, et al., 2013). One of these systems is the Mercedes Distronic Plus with steering assist, which is a so-called level 2 system (explained below), allowing the driver to set a speed that the vehicle follows. This system also helps the driver with braking and throttle control, as well as lateral lane guidance by steering back when the vehicle drifts out of the lane (Mercedes-Benz, 2015).

1.1.2 *Levels of Automation*

There are different levels of automation, from barely any system assistance to a fully autonomous system (Marinik, et al., 2014). One organization that has presented one of the more accepted taxonomies is the National Highway Traffic Safety Administration (NHTSA) under the U.S Department of Transportation. NHTSA is working for a more safe traffic environment in the U.S. and their definition of levels of automation is a way of establishing a universal language in order to understand the basis of these systems. The different levels range from no automation, defined as level zero, to fully automated and self-driving systems, designated level four (Marinik, et al., 2014).

Levels of Automation (NHTSA, 2013)

- **Level 0 (no automation)**
  The driver has all control.
- **Level 1**
  (function - specific automation)
  Assisting functions, e.g. braking.
- **Level 2**
  (combined function automation)
  Uses at least two automated functions i.e. adaptive cruise control together with lane centering.
- **Level 3**
  (limited self-driving automation)
  The vehicle can drive on its own within a certain context e.g. highways. The driver does however need to be available for occasional control handovers.
- **Level 4 (full self-driving automation)**
  The vehicle can take the driver from point A to point B solely on its own without driver participation needed except for providing a destination.

Something to be noted is that NHTSA´s levels only work as guidelines for companies and legislators and not as regulations (KPMG, 2013). Level 3 in NHTSA, which is the second highest automation level, will be the level on which this thesis work will focus. Level 3 in NHTSA, called limited self-driving automation, states that the automation should be able to take control over all safety-critical functions, hence have all driving control over the vehicle within a certain context such as highways. The driver should however always be ready to reclaim control due to a change of context, e.g. in urban areas or similar high complexity conditions (NHTSA, 2013).

1.2 Background

1.2.1 *Safety & Human Failure*

Autonomous driving vehicles could help lower the number of accidents within the infrastructure of the road network according to Leohold (2011). This is because so-called human error accounts for 95 percent of all fatal accidents and technology only for five percent. This is also supported by Gasser, et al. (2013), who argue that implementing highly autonomous driving vehicles and systems will lead to a decrease in “human failure” and, by that, significantly lower the number of accidents. There are predictions about an increase in road capacity by 500 percent due to autonomous vehicles systems (KPMG, 2013). But even if the systems become flawless, it is the demands and requirements from the market that will have the final saying in whether or not these highly autonomous driving vehicles will succeed as a consumer technology (KPMG, 2013).

1.2.2 *Trust Issues*

As stated above, the technology is already here for autonomous vehicles, but even if the technology already exists, there are still problematic issues that have to be addressed. One specific issue is related to human interaction, namely trust. This can be regarded as one of the biggest issues within HMI in these types of vehicles according to Verberne & Ham (2012). This is because it is crucial to have trust in order to create acceptability for the systems (Verberne, et al., 2012), which is a precondition for using the systems. Trust is needed for AD vehicles because there are certain risks, uncertainties, as well as a mutual dependency between the driver and the vehicle (Mcknight & Chervany, 2000) (Muir & Moray, 1996), and without trust people will probably not use the vehicle even if the AD system’s driving performance is good.

1.2.3 *Project Aim*

The aim of the thesis is how to create an appropriate level of user trust for future autonomous driving (AD) vehicles. In order to do so, a greater understanding of what affects trust in the human-machine interaction in autonomous driving vehicles is explored. This knowledge is then used to generate a guiding framework for implementing trust-related factors into a human-machine interaction system. Furthermore, a human-machine interaction system example concept is designed that could mediate an appropriate level of trust according to the system’s actual performance, and by that, enable the driver to use the autonomous driving system correctly.

In order to accomplish this, the following main research questions will be addressed:

What are the factors affecting trust in a human-machine interaction, specifically in the context of autonomous driving vehicle systems?

What events during the human-machine interaction in an autonomous driving vehicle are affected by these factors?

How can a trust-based framework contribute to designing trust into HMI systems in AD vehicles?

1.2.4 *Project Goal*

The project goal is to identify factors affecting trust between the human and the machine. These trust factors will then be placed into a trust-based framework formulated around a driving scenario, which explains different events taking place during the interaction between the driver and the autonomous driving system from a trust perspective. A second goal is to use this framework to produce HMI concepts of which a specific feature will be evaluated through a validation test in order to validate which of the concepts creates the most appropriate level of trust for the driver. The framework should guide future human-machine interaction designers in building trust into the human-machine interaction system.

1.2.5 *Delimitations*

This thesis will only regard NHTSA´s automation level 3, Limited Self-Driving Automation, because this level covers more or less all parts of automation, but is narrowed through contextual limitations e.g. highways (Marinik, et al., 2014).

Only available autonomous driving vehicle systems of today or systems within the near future, i.e. five to seven years ahead, will be considered when designing and developing the concepts.

The designed HMI concepts only consider internal status communication, i.e. not external circumstances such as other intelligent systems or groups, since the thesis' focus lies on the driver's trust for the system. Only one HMI feature of the developed concepts will be tested due to time limitations.

Due to legal issues, user study participants were limited to people connected to Volvo. Thus the participants all have engineering backgrounds, which limits generalization.

Since there were not any level 3 autonomous vehicles available during the user study, a level 2 vehicle had to be used instead. This was also an issue during the validation test, in which a few concepts were evaluated, where a right-hand driven vehicle had to be used.

1.2.6 *Definitions*

The definitions presented here are rather explanations over certain abbreviations as well as clarifications how certain words are used in the thesis.

Autonomous: *Refers to a machine that is automated and by this can perform tasks more or less on its own.*

Autonomous Driving System(s) (AD System): *A system that handles the vehicle through automated functions.*

Autonomous Driving Vehicle (AD Vehicle): *A self-driving vehicle that uses AD systems and through this is autonomous.*

Driver Information Module (DIM): *The display that is placed in the dashboard in front of the steering wheel and shows vehicle information such as speed (speedometer) and engine revolutions (tachometer), etc.*

Center Stack Device (CSD): *The display placed in the center of the dashboard that often contains the infotainment systems as well as other information and functions.*

1.2.7 *Distribution of Work*

The authors, Fredrick Ekman and Mikael Johansson, conducted the project work equally.

1.2.8 *Social Sustainability*

Social sustainability is defined as *"a positive condition within communities, and a process within communities that can achieve that condition"* (McKenzie, 2004, p. 23)

With this definition in mind and with the recent research regarding autonomous vehicles in hand, the introduction of these could entail a positive outcome from a social sustainability perspective. This, since these types of new vehicles and systems could reduce the number of accidents and at the same time create an increase in road capacity, allowing more people to be able to use the already existing road network. Social sustainability is all about creating healthy and livable communities, something that the progression of autonomous driving vehicles could contribute to since it could decrease the number of accidents (McKenzie, 2004).

The introduction of autonomous vehicle could also alter the transportation paradigm as it is today since it could facilitate greater mobility without having the same car ownership as we have today, by allowing more car sharing solutions that could lead to a more energy efficient use of cars (Sustainable Mobility, 2013).

By aiming for a shared transportation community it could allow more people to carpool and at the same time reduce energy consumption. But if this scenario ever will see daylight it is important to create a milieu that is desired for the user for instance through allowing customizable or personal profiles within the vehicle creating a shared platform for all.

It is also possible that the introduction of autonomous vehicles could enrich everyday life for people through letting the driver use the time usually spent maneuvering the vehicle in a way that is more convenient.

2 Literature Study

This chapter presents the information gathered from the literature study regarding human-machine interaction and trust. The methods used in the literature study as well as in the rest of the project are described in the following chapter, 3 Methodology & Implementation.

2.1 HMI – Human-Machine Interaction

As the driving situation is changing rapidly and moving towards more automation, the driver's role will change from being an operator to becoming more of a system supervisor (Merat & Jamson, 2009). This is going to result in new demands on the driver as well as on the collaboration between the human and machine. The complexity of these automated machines are getting more and more sophisticated and "intelligent" and, as humans, we need to contemplate how much of this complexity we, as "drivers", need and want to see when operating a vehicle, since this could have a great impact on our behavior (Thill, et al., 2014).

There are several issues regarding autonomous driving vehicles that have to be solved in order to be able to fully implement self-driving systems in everyday life. In aviation where these systems already are in use, both positive and negative effects on the operator have been identified (van Schijndel-de Nooij, et al., 2010).

For driving, the main human factors concerning human-machine interaction (HMI) can be classified as safety, usability and acceptance (Saffarian, et al., 2012). These factors can be further divided into subgroup issues such as over-and under reliance because of an insufficient mental model (Saffarian, et al., 2012), loss of situation awareness, loss of skill as well as quick changes in mental workload or too much or too little workload (Merat & Jamson, 2009) (Toffetti, et al., 2009). Some of these issues are because the systems may keep the human "out of the loop" when performing a task, and do not give sufficient feedback, which means that it can be hard to differentiate between which tasks the systems perform and what the human user, in this case the driver, should be responsible for (Hoc, 2000). This can cause the user to not act when supposed to or act in a less favorable way; this phenomenon is called mode-confusion. It is therefore important to understand the automation's limitations because these automated functions will not operate as a human would act and cannot therefore be seen as one (Saffarian, et al., 2012).

Training and well-designed interfaces such as different types of displays could optimize driver performance, could help to prepare the driver for possible errors (Wickens, et al., 2010) and allow a preferable level of trust to form. It could also be especially important to use certain reminders such as different types of signals in order to inform the driver about upcoming events and situations; thus keeping the human operator in the loop (Merat & Jamson, 2009). These areas need to be further investigated and considered in order to optimize the human – machine interaction in autonomous vehicles; not only the technical systems per se but how the technical systems will function and be designed in a manner that creates trust for the users.

In the literature, six major HMI issues have been identified which are closely related to each other and affect one another in complex ways. These factors are situation awareness, mode confusion, usability, loss of skill, workload and trust. Since problems correlated to trust are the main focus of this thesis, trust has been separated from the other HMI issues into a section of its own.

2.1.1 *SA – Situation Awareness*

Situation awareness (SA) can be defined as understanding what is happening around you, or more formally as *"the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future"* (Endsley, 1995, p.36) meaning situation awareness can be seen as a model regarding the situation and setting around oneself. Situation awareness, hereafter referred to as SA, is an important part of achieving a safe driving environment in autonomous driving vehicles (Koo, et al., 2014) and needs therefore to be considered. SA is mostly important in rapidly changing situations were focus is needed.

SA can be divided into smaller pieces in order to easier grasp the concept, namely, perception, comprehension and projection (Thill, et al., 2014). Perception regards the driver's ability to receive input from the surroundings, both internal, e.g. the system (also known as system awareness) as well as the external surroundings as the environment outside the vehicle. Comprehension is basically how the operator understands the current input from the surroundings and projection is the estimation how these surroundings will change within the near future based on the current situation.

The main problematic issue regarding SA in autonomous vehicles is when the autonomous system is in control but a sudden change occurs which will shift the control back to driver (Parasuraman, et al., 2008). Here the driver suddenly needs to have a great understanding of what is happening and how the environment looks in order to take over control in a safe way, hence a high level of SA is needed. The aviation industry has also seen low or insufficient SA as one of the biggest issues regarding human factors in autonomous systems (Stanton & Young, 2000).

Automation itself without good feedback or information output could have a negative impact on SA since it can cause an overreliance on the system (Merat, et al., 2012). This overreliance could cause human users to hand over too much control to the system, more than the system is designed to handle, which can cause a critical situation. It is therefore crucial for the driver to know the system's capabilities, what the surroundings look like and how the context is changing in order to quickly make a control transition from AD-mode to manual mode, i.e. human control (Merat, et al., 2012). As such, SA is desirable in order to have effective human-machine interaction and by that, safe driving conditions.

As stated above it is of major importance as a driver to have a good perception and comprehension of the current context as well as have an understanding of future upcoming events (projection) to keep the driver in the loop (Merat, et al., 2012). This understanding of context and changes within this context is fundamental to good SA. In order to enhance SA it is important to assist the drivers to know where to find the right information about the system and its intentions rather than trying to teach the users what to do during certain scenarios (Parasuraman, et al., 2008). The system itself could be the source of helping the users to get the right information by giving correct feedback, such as contextual information as well as system information, at the right time as well as helping the driver to integrate it and by this understand the information both in the current context as well as its future implications. Since SA involves spatial, temporal, goal and system awareness the feedback and information given from the system should be given through comprehensive and clear displays (Thill, et al., 2014) making it easy to understand as well as showing its intentions about what it is doing and what it is going to do in order to keep the driver in the loop and by this creating an optimal level of SA.

2.1.2 *Mode Confusion*

Mode confusion can be seen as to poor understanding about which level of automation is active and which functions and limitations it has. This confusion often occurs at control transitions between different driving modes, for instance, AD-mode to manual mode and vice versa, when the driver does not know what is expected of him or her. It is also a result of not knowing what the system's intentions are and what it is capable of. The problems connected to mode confusion most often arise when the driver needs to take control over the vehicle for instance in a critical situation when the system lacks the ability to solve the situation with a preferred outcome (Toffetti, et al., 2009). Three main questions have been identified that need to be contemplated in order to grasp the issues regarding mode confusion, namely, "*Who has the control and to what degree? Who should get the control and to what degree? As well as who initiates the transition?"* (Toffetti, et al., 2009, p.5).

The role of the driver will change with increasing levels of automation and the driving tasks will become more dynamic, moving away from driving and more toward monitory tasks, creating more control transitions. This creates a greater need for the system to convey the correct information about the system status, since an issue with inappropriate actions from the driver could lead to mode error, i.e. if the driver does not fully understand which automation mode that is active, to what extent the automation is engaged and what it does, this could lead to the driver performing a less preferable action for that specific situation (Saffarian, et al., 2012). In these highly automated modes there is a need for a full understanding of the system's intentions, mode specific tasks and automation level, since when drivers in the future will perform other tasks during AD-driving, it could lead to complicated scenarios if system failure arises and the driver needs to take control (Merat, et al., 2012). It is therefore also important that the driver gets enough time to be able to take in and comprehend the mode information so that transitions from for example AD-mode to manual drive will be safe and controlled. The system should also be "intelligent" so that the need for a control transition is communicated from the system to the driver (Gasser, et al., 2013).

In order to solve these issues the interaction and interface design is the key factor because a well-designed HMI design could raise mode understanding and help the driver to monitor the automation. This interaction design through some type of interface should primarily focus on the levels of automation and especially the transitions between the different levels (van Schijndel-de Nooij, et al., 2010).

2.1.3 *Usability*

The usability of an autonomous vehicle system is directly connected to the transparency of the system as well as to its simplicity and accessibility, which is always important, but even more so in critical situations (Dekker & Woods, 2002).

It is important to design clear and intuitive displays in order to create an HMI system that is effective (Hancock, et al., 2013) as well as making it easy to use or focus on training the driver in order to create an expert understanding of the system characteristics (Wickens, et al., 2010) (Lee & See, 2004). The information presented by the system through the interface should be presented in a way that is easy for the user to comprehend. An important factor to achieve this is to have concrete and preferably detailed information with a constant and structured appearance (Lee & See, 2004). It is also of major importance that the information is presented in a correct cognitive way in order to allow the driver to fully understand the information received from the system as well as keeping it well-balanced i.e. not too much nor too little information (Davidsson & Alm, 2009). Factors such as simplicity, balance, intuitiveness, and structured and detailed information should be considered when designing an human-machine interaction system that is optimal for the user and in this case the driver.

2.1.4 *Loss of Skill*

Long-term use of autonomous vehicles, especially NHTSA levels 3 and 4, could lead to unwelcome changes in the operator's manual driving skills because of always relying on the automation, a phenomena commonly known as loss of skill (Stanton & Young, 2000). Loss of skill can be divided into two different categories: long-term and short-term loss of skill. The difference between them is in which time frame they affect the loss of skill, where short-term loss presents immediately and long-term loss occurs after a longer time-period of automation usage (Stanton & Young, 2000).

Short-term loss of skill appears when a transition is made between two levels of automation and the driver cannot cognitively cope with the new situation fast enough. Long-term loss of skill on the other hand is due to a decreased level of manual driving which leads to lessened manual driving ability. This will probably only come into play in the autonomous vehicle sector when automation has come so far that human drivers more or less only rely on the automation itself.

The problematic issues concerning loss of skill is that it could lead to complacency, trust and self-confidence issues as well as adaptive concerns in unforeseen situations when automation itself is not competent enough to solve the problem at hand and human intervention is needed (Hoc, 2000). This also leads to an accountability problem, i.e. who is the one in charge if the human operator is more incapable than the system to solve different tasks. It is possible to say that loss of skill is due to the introduction of a high technical system that leads to a more passive role for the human operator.

The loss of important manual control skills is still a major problem within the aviation sector leading to less cognitive and performance abilities (Saffarian, et al., 2012). Possible solutions could be to either lower the levels of automation, which may not be a favorable way of solving the problem, or to design HMI systems that demand less expertise or are at least highly understandable, which will not solve the problem but could minimize the impacts of the loss of skill (Wickens, et al., 2010). This is because if the system is simpler to use, the loss of skill will play a smaller role to the outcome of the situation.

2.1.5 *Workload*

Workload can be seen as the relation between the mental work it takes in order to do a task and the mental resources of a driver (Parasuraman, et al., 2008). The amount of workload a driver perceives is also connected with that person's ability to comprehend what is happening around him or her (Hoff & Bashir, 2014).

Automation will, in ordinary scenarios without any critical situations, likely lower the driver's mental workload since the driver no longer needs to be as active in the act of driving, especially in routine operations. However, a higher degree of automation could also increase the workload for the driver if the system needs a lot of supervisory controlling (Wickens, et al., 2010).

The biggest mental workload related issues arise when a sudden change of needed workload appears and the driver must handle a greater mental workload in a short amount of time. There are results showing that reduced mental workload could lead to an unsafe driving situation when system failure or an unforeseen situation occurs and the driver needs to take over the control, but since the driver is out of the loop it could be difficult to respond in a preferable way (Merat, et al., 2012).

The reduction of mental workload goes hand in hand with lowered SA and is a side effect of higher levels of automation. However, mental workload can also increase significantly during unexpected critical driving situations (Saffarian, et al., 2012) (Hancock, et al., 2013) (Parasuraman, et al., 2008), causing the driver to not be able to meet the needed control transition properties that are needed.

There are solutions that could minimize the problems around too high or too low a workload. Interfaces could be designed in such a manner that allow the user, in this case the driver, to get the information at the right time in order to be able too fully evaluate the information presented by the system (Parasuraman, et al., 2008). This will not directly solve workload issues, but it will give the driver time to adapt to new level of workload. The information that is presented should also be presented in a way that is accessible and understandable since easily understandable information can lower the mental workload needed for the driver to take in the information. In order to still be able to process the great amount of information presented, it is then needed to use simple, clear and understandable information cues, preferably familiar pattern-based cues, such as icons and imagery which help to minimize the cognitive workload (Dekker & Woods, 2002).

2.2 Trust

The section about trust is divided into three subsections: fundamentals, issues and factors. In 2.2.1 Fundamentals, the concept of trust will be defined and the fundamentals of trust formation will be presented. Subsection 2.2.2 Issues will introduce the problems and issues connected to trust in automation, and in the subsection 2.2.3 Factors, a number of trust factors will be presented that could help solve these issues.

2.2.1 *Fundamentals*

Here the fundamentals of trust, both interpersonal and human-to-automation, will be introduced to lay a foundation for the rest of the thesis. It contains seven parts, Definition, Components for Trust to Establish, Trust as an Attitude, Trust Formation Order, Basis of Trust, Cognitive Processing and Contextual Components.

Definition

Trust is a necessary and valuable factor whenever there is risk, uncertainties or interdependencies (Mcknight & Chervany, 2000) (Muir & Moray, 1996) and is perhaps one of the most important cornerstones in autonomous vehicles in order to create something that will be a positive experience for the user (Waytz, et al., 2014) (KPMG, 2013). The concept of trust can be hard to define and different disciplines such as psychology, sociology, economics and human factors look at trust in different but similar ways and have their own definitions (Mcknight & Chervany, 2000). When we talk about trust, most people have an intuitive understanding about the concept of trusting someone. People may have a hard time pinpointing what it really means, and they can probably mention a few characteristics that they connect to the expression of trust, although it may vary a lot from person to person. What we often mean when we talk about trusting someone is that the chance that the person will perform a certain task is so high that we are willing to interact with him or her (Fishmana & Khanna, 1998). Lee & See (2004) have defined the concept of trust in a more formal way as “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability”.

Components for Establishing Trust

For trust to even be needed and begin to establish, certain criteria need to be present. Hoff & Chervany (2014) have presented three components which are needed; first there needs to be a trust giver and a trustee, an incentive as well as a possibility for the trust giver to fail (see figure 1). This means that in order for trust to establish, there needs to be an agent to give or instill trust (trust-giver) and an agent to receive or experience it (trustee). There also needs to be an incentive for the trustee to perform the task, which in the case of automation often is the designers’ intended use. The third component states that there also needs to be a risk for trust to be needed.

Figure 1 - Components for trust to establish.

Trust as an attitude

Just because the agent is trusted does not mean the agent is reliable since there are other factors influencing the decision. This can be explained with Ajzen & Fishbein's (1980) framework presented by Lee & See (2004) which tries to explain how a belief can turn into a behavior. This framework consists of four hierarchical steps, where the outcome of each step influences the next one. It starts with beliefs forming an attitude that creates an intention which may lead to a behavior (see figure 2). In this framework trust is considered an attitude that is formed by information about and impressions of the system. Depending on the information and impression, different levels of trust will be attained leading to different intentions to rely on the agent. This intention to rely on the agent may then turn into a behavior but there are also other factors such as HMI issues that affects if the intention turns into a behavior.

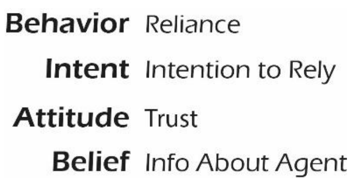


Figure 2 - Framework showing how a belief is turned into a behavior by Ajzen & Fishbein (1980)

It is therefore important to view trust as an attitude and not a as a behavior since there are other factors such as workload, situation awareness and self-confidence that can influence the trustee's behavior (Lee & See, 2004).

This is why the HMI factors are important to take into account when developing a HMI system with regards to trust. (Mayer, et al., 1995).

Trust Formation Order

Human-automation trust can be seen as a sort of interpersonal trust and the supervisory control of automation shares a lot of resemblance with the interaction between staff and manager (Muir, 1994) (Hoff & Bashir, 2014). Concepts about trust between human and human are often similar to trust between human and automation even though it seems like people consider the human-to-human interaction more in terms of trust rather than distrust (Jian, et al., 1998). One possible explanation to why the concepts are similar is that trust in automation may reflect the trust in the designers of the automated system (Muir, 1994), i.e. it may be that we do not actually think about the system per se when we talk about trust, instead we think about the person creating the system, putting our trust in them.

Muir (1987) presented a hierarchical three-stage model by Barber (1983), how trust is formed and evolves, were each stage is depending on the outcome of the previous stage. The three stages are predictability, dependability and faith (see figure 3). This means that in the beginning of a relationship we often base our trust in another person on how well we can predict their actions and later we base our trust on the dependability of another person's actions, which can be seen as summary of the actions of the trustee. The final part of trust development occurs in a fully mature relationship and here trust is based on faith that the person will continue performing in the same way.

In contrast to interpersonal trust, trust between human and automation often progresses in the opposite order, where faith is important early in the relationship, followed by dependability and predictability (see figure 3) (Lee & See, 2004). This is because people often tend to be positively biased in their trust in automated systems because automation is looked upon as something that is expected to perform better than the human counterpart (Dzindolet, et al., 2003). Another explanation to the fact that humans often tend to be biased towards automated systems is that trust may be based on the reputation of the automation, which could be inaccurate (Hoff & Bashir, 2014).

Figure 3 - Hierarchical three-stage model showing trust formation order. First interpersonal trust (Barber, 1983) and then trust between Human and Automation (Lee & See, 2004).

Basis of Trust

In interpersonal trust there are three characteristics that have been identified as especially important in order for trust to grow: Ability/competence, benevolence and integrity (Mcknight & Chervany, 2000) (Mayer, et al., 1995). These three characteristics form the base on which trust grows.

Since interpersonal trust and trust between human and automation are very similar, these three characteristics have three corresponding characteristics in trust between humans and automation. Here these characteristics of trust are referred to as performance, purpose and process (Lee & See, 2004). These characteristics both for interpersonal trust as well as for trust between human and automation can be seen in figure 4.

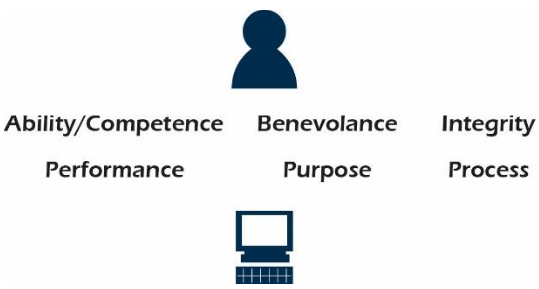


Figure 4 – Important characteristics for trust, both interpersonal trust (Mcknight & Chervany, 2000) (Mayer, et al., 1995) as well as trust between human and automation (Lee & See 2004).

Purpose-based trust is related to the designer's intended use of the automation, which describes why the automated system is developed. Purpose is connected to benevolence but since the automation to date does not have an intention of its own, it is therefore the designer's intention that is facilitated into the system. Process-based trust correlates to the interpersonal basis of integrity and reflects the user's understanding of the system, since a user often trusts a system which can be understood and that is perceived as able to achieve the user's goals (Lee & See, 2004). Performance-based trust is similar to the Ability/competence characteristic and is affected by the operations of the automation, and therefore the trust will vary depending on how well the automation performs (Hoff & Bashir, 2014). Trust can be formed "…from a direct observation of system behavior (performance), an understanding of the underlying mechanisms (process), or from the intended use of the system (purpose)" (Lee & See, 2004, p.67). If trust is based on several of these factors it will become more stable and robust then if it is just based on a single one of them (Lee & See, 2004). This means that if the user's trust is both based on the performance of the system and the understanding of the system's intention, the trust will be more stable than if it is only based on the system's performance.

Cognitive Processing

How the users perceive the information they receive is of great importance since trust is based not on the actual trustworthiness of the system but on the user's perception of it. It is therefore important to know how the user processes the information presented.

The information affecting trust is processed in an interplay between three different cognitive processes: an analytical, an analogical and an affective process (see figure 5). Analytical processes can be seen as reasoning based on existing knowledge and information about the system. This is the most cognitively demanding process since the information received is consciously analyzed but it is also the type of process that will give the most appropriate level of trust (Jenkins, et al., 2015). The analogical process uses solutions of old problems to create rules or procedures, which are compared with the system to assess the trustworthiness, but these rules and procedures can also be affected by gossip and reputation (Lee & See, 2004). The last cognitive process, the affective process, is considered the most important and is affected by emotional impressions and feelings (Lee & See, 2004). Depending on the situation, the user will have different levels of cognitive resources available, which will affect which cognitive processes are used since they have different demands. For example, if the system creates a situation where the driver has a low level of cognitive resources available, he or she will not be able to process information within the analytical process.

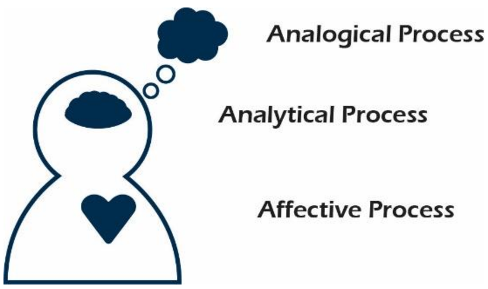


Figure 5 - Cognitive Processes.

Contextual Components

When trust is formed there are a lot of different variables involved in the formation process that can be seen as contextual components affecting the trustee. Trust formation can be seen as dynamic interaction between three contextual components; the operator, the environment and the automation (Lee & See, 2004). Marsh and Dibben (2003) identified three layers that correlate to the three contextual components and they consist of dispositional trust, situational trust and learned trust (see figure 6).

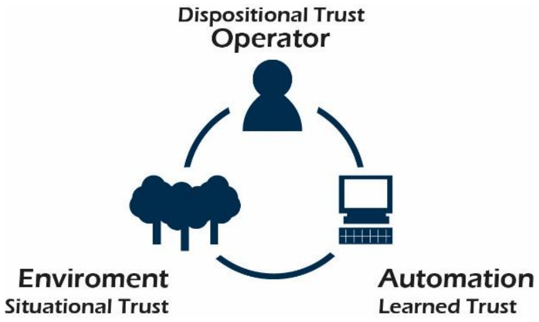


Figure 6 - Contextual Components (Marsh & Dibben, 2003).

Dispositional trust that correlates to the user component is characterized by individual willingness of the user to trust an agent, which is quite stable over time compared to the two other components (Hoff & Bashir, 2014) since personal traits have a tendency to remain the same. Four major factors affecting dispositional trust are culture, gender, age and personality (Hoff & Bashir, 2014). Research has shown that there are cultural differences affecting the individual's tendency to trust, and statistics suggest that western countries have a higher mean level of trust compared to all other cultures (Fishmana & Khanna, 1998). It has also been identified that age plays a large role when it comes to the tendency to trust. For example studies have shown that older adults have a higher tendency to trust automation in early parts of the interaction than younger adults but the trust evolves the same way later on (Hoff & Bashir, 2014). Even though specific differences between genders have not been established, some findings have shown that males and females respond differently to different communication styles , therefore making gender an important factor (Hoff & Bashir, 2014).

In the personality factor, several sub factors affect the tendency to trust, but one that has been identified as especially important is a person's locus of control (Helldin, et al., 2013). Locus of control describes people's propensity to blame external or internal factors when events occur. With a high internal locus of control, people tend to believe that things that happen are their own fault, while people with a high external locus of control have a hard time accepting blame but instead believe in environmental reasons. It has been shown in driving situations that high internal locus of control has been related to alertness and self-bias in accidents, while high external locus of control has been associated with aggression and tenseness (Stanton & Young, 2000). Locus of control can not only be seen as a factor affecting people's tendency to blame themselves in driving situations but also the tendency to blame automation. This is why locus of control needs to be taken into consideration when talking about trust.

Situational trust is more rapidly changing than dispositional trust since the environment is affecting this layer of trust. Situational trust is context specific and depends both on the external environment as well as the internal environment, which is the user's characteristics that change according to the environment such as mood, expertise in the subject, self-confidence and attentional capacity (Hoff & Bashir, 2014). The external factors consist of the system's complexity and task, the risks and benefits of the system, the operator's workload and the organizational context, which are the effects of reputation, gossip, formal and informal roles before the user has had direct contact with the system (Lee & See, 2004).

The last layer, learned trust, is formed by the user's perception of the system and can be divided into two parts, initial and dynamic, where initial learned trust is the preexisting knowledge before interaction and dynamic learned trust is the trust created during the interaction (Hoff & Bashir, 2014). It has been found that users trusted systems more after gaining experience with the system and also that the trust was more and more influenced by the system's characteristics and less by individual tendency to trust the automation (Merrit & Ilgen, 2008). This means that a user's individual tendencies and traits play less of a role the longer the interaction with the system goes on and the more experience is gained. Learned trust can be seen as a circular interaction where the experience with the system and how it performs will affect the level of trust leading to new use patterns. If the performance of the system would alter the level of learned trust this may lead to the user altering his or her way of using the system, which could in turn affect the performance of the system which could lead to a vicious cycle (Hoff & Bashir, 2014). This shows how complex the interaction between the system and the user's trust is. It is also important to remember that design features will not alter learned trust per se but will change the way that the user thinks the system performs (Hoff & Bashir, 2014).

2.2.2 *Issues*

Over- and under reliance

The main issue of forming trust in automation is over- and under reliance on the system (Szymaszek, 2014). This occurs when the human operator overly trusts the automation system, which could lead to misuse, or when the human operator has lower trust for a system that is more competent than the operator, which may lead to disuse (Dzindolet, et al., 2003). Overreliance has been identified as a major factor in several commercial airplane crashes and under reliance is believed to have played a significant role in the Costa Concordia disaster (Hoff & Bashir, 2014). The optimal trust level is therefore not a as high trust level as possible but the level that correctly reflects the automation's actual competence level (Merrit & Ilgen, 2008).

Studies have shown that if an operator has a highly reliable automated system and at the same time has other tasks to perform, the operator will probably allocate the attention to the other tasks (Parasuraman, et al., 2008). This phenomena, called the allocation strategy, can itself be a result of a too high level of trust and may lead to less observant monitoring, which could cause the operator to miss automation failures leading to even higher levels of trust (Lee & See, 2004). This spiral of possibly higher and higher trust suggests that it may be hard to achieve a proper level of trust if the driver has a lot of secondary tasks beyond the automated system itself.

It has also been found that experience and practice alone are not sufficient in counteracting overreliance since the effect is still present in experienced pilots and controllers, and it has been shown that additional task training in naive operators does not eliminate overreliance either (Parasuraman & Manzey, 2010).

These findings point out the importance that the user has a well-calibrated level of trust to not experience over- or under reliance on the system. This is mostly done through "improving the accuracy of the operators´ perception of machine competence" (Muir & Moray, 1996, p. 454) which can easier be achieved if users have a good understanding of the system and their own attitudes toward it (Muir & Moray, 1996). So it is potentially achieved by giving the user a more appropriate picture of how the system works and its purpose. If the system's functions and reliability vary over time, it is more likely that the user will try to adjust the level of trust to the specific situation, creating a more appropriate level of trust (Parasuraman & Manzey, 2010) (Lee & See, 2004).

Understanding

To be able to build a proper level of trust it is important for the user to have a good understanding of the system and how it makes its decisions. If the users fully understand the system or if the system is used in a context or environment that is very familiar, trust will play less of a role on the reliance (Hoff & Bashir, 2014). It is therefore of great importance that the users have a correct understanding of how the automation makes it decisions. If the users have too little understanding of the system they may see it as untrustworthy and reject it or they may not be aware of the system's limitations. One example stated by Saffarian et al. (2012) is that adaptive cruise control systems have the ability to maintain constant speed but the sensors have limitations and if the users are not aware of the limitations, they may not be able to reclaim control in case of an emergency. It is therefore important that the users have a good understanding of the limitations or get information about system limitations well in time of having to intervene.

Transparency

In order to obtain system understanding the user needs to receive accurate feedback and information about the automation. It may be especially important in a highly automated system, where handling errors are more difficult, to be presented with relevant information of whether the system works or not (Toffetti, et al., 2009). A transparent system will give the user an increased feeling of control since the user has a greater ability to predict the system's behavior (Verberne, et al., 2012) which gives a sense of knowing what the system is "thinking". Since much of the trust is based on observations of the behavior, it is a crucial factor that the behavior is made observable (Muir, 1987). If the users are not aware of what the system does they will not be able to get a proper level of trust since they will not achieve a correct understanding of the system.

Feedback

To achieve transparency in the system it is not only important that the user gets feedback but it is also very important what feedback is presented and how. The feedback must be clear enough for the user to notice it but at the same time it cannot be intrusive since it may annoy the user (Saffarian, et al., 2012) or even be distracting for the user (Stanton & Young, 2000). In a car it may be especially important not to have too much information since the visual inputs are already quite significant (Davidsson & Alm, 2009). This can be especially hard since different users can have different preferences and find different types of feedback desirable.

Timing of feedback is another important factor, as too early feedback or feedback presented at the wrong time (such as false alarms) can lead to distrust and in the worst case, the user might even turn the system off (Saffarian, et al., 2012). So for the feedback to be favorable it needs to be presented in a timely way, not too late and not too early, and it should be prominent but at the same time not too intrusive, and at the same time fit most people. This implies that it in some way needs to be adaptive or customizable to be able to cope with all demands. The feedback should also be suitable for all three types of cognitive processing not only the analytical and analogical since the affective (emotional) process is the most influential on trust (Lee & See, 2004).

Mental model

Creating a proper level of trust is very complex since "trust is more than a simple reflection of the performance of the automation; appropriate trust depends on the operator's understanding of how the context affects the capability of the automation" (Lee & See, 2004, p.72). To be able to comprehend this complex interaction between the user, environment and system, the user forms sort of a "picture" of the situation called a mental model. Mental models help the user predict and explain the interaction but they are only an approximate representation and can often be "incomplete, unstable, ad hoc and unscientific, even superstitious" (Stanton & Young, 2000, p. 325). How the user's mental model looks affects how they use the system and in some cases the mental model does not match the observed behavior of the system, leading to confusion regarding the capabilities of the system and who's in control (Toffetti, et al., 2009). Knowing which mental model a user has or how to create a new one is not easy since humans are rather complex, but taking into account how the interaction with the system can affect the user's mental model is critical when designing the system (Stanton & Young, 2000). It is therefore important that the involved individuals in the design team have a coherent vision of the system to be able to form the user's mental model in good way.

Loss of Trust

Trust takes time to build up but is fragile and will more rapidly decrease in case of errors and then slowly build up again, under the right circumstances (Muir & Moray, 1996). This phenomenon may occur due to the fact that information that does not correlate to expectations is often well-remembered, having too big an influence on trust formation (Dzindolet, et al., 2003), giving the user the wrong image about the system's trustworthiness. This makes the first impression and early use especially important since an early error can have long-lasting effects and can be hard to recoup from (Hoff & Bashir, 2014). It has also shown that if the performance of the trustee does not match the intended purpose, it will have substantial effects on the loss of trust (Lee & See, 2004). This shows how crucial it is for the system to express the same purpose as it is able to perform.

Research has shown that errors in performance significantly affect the user's trust in a negative way even when the outcome of the errors are temporary or not affecting the performance in the long run (Muir & Moray, 1996). One explanation of this could be that a lot of the automation we encounter every day, for example calculators, works either perfectly or not at all, and in these cases an error means that the automation is faulty and not to be trusted (Dzindolet, et al., 2003). Another explanation to why even temporary and non-critical errors affect the user's trust in such a significant way could be that the user may think that if the system cannot even handle a basic function then it will not be able to handle more complex situations (Hoff & Bashir, 2014). One way to counteract this loss of trust could be to explain that the error could occur without affecting the outcome in a negative way (Dzindolet, et al., 2003). This understanding could lead to a reduction in the trust decline by giving the user a higher degree of system understanding. Muir & Moray (1996) found in their studies that if the errors occurred in a consistent manner the user could find ways to work around the problem by knowing when it would occur and the trust started to grow again. This was not the case when the errors occurred in a variable way. This shows that the predictability of the errors has more impact on trust than the severity of the error. Stanton and Young (2000) have also argued that distrust only spreads inside subsystems and does not affect the trust for other subsystems or the whole system, possibly leading the user to shut off a subsystem because of distrust but at the same time trusting and using the other subsystems. If the subsystems are not able to be disconnected individually this may force the user to shut off the whole system because of a too low level of trust for one part of the system, probably leading to disuse.

There is a better chance for the user to build a proper level of trust if the level of trust is high in the beginning since if it is low in the beginning it is possible that the user will not use the system, leading to an even lower level of trust and use, causing a downward spiral of the trust level (Lee & See, 2004).

Not all issues affecting the reliance behavior are affecting the user's intention to trust. Hoff and Bashir (2014) have identified different factors affecting the user's reliance on the system without necessarily affecting the trust. The factors are; time constraint, effort level to engage the system, alternatives to using the system, user's situation awarness, workload and physical well-being.

If trust issues as the ones presented are not considered in the HMI designing process it could jeopardize or at least slow down the evolution within autonomous vehicles, since it could create a mistrust for this type of fairly new technology.

2.2.3 *Factors*

In the literature several factors were found that could help solve or at least counteract the issues with trust between automation and humans in order to create a proper level of trust. The factors are feedback, error information, uncertainty information, why & how information, training, two-way communication, common goals, adaptive automation, anthropomorphism, customi-zation, expert/reputable and mental model.

Feedback

An especially important factor is feedback or more specifically how feedback is presented. The feedback can be split into two different categories, action feedback and learning feedback (Stanton & Young, 2000). Action feedback is where the user gets feedback whether the action was successful or not immediately after an action is performed. This type of feedback supports fast learning but also fast skill degradation when feedback is removed (Stanton & Young, 2000). Learning feedback, which is often given during training, consists of more detailed feedback about the performance, which leads to slower learning but also more persistent skill knowledge (Stanton & Young, 2000). A combination of these two feedback styles would be preferable to enable long-lasting skill knowledge but to also get the fast learning so the user gets a direct understanding of how the system works. The user should get the feedback and information about the system's actions continually to be kept up to date and be more prepared in case of a takeover situation (Toffetti, et al., 2009). Giving continual information provides the user with information about the predictability of the system and also an insight into the process. It has been found that providing information in transitions of control between user and automation is important. Automated aids that provided information when taking over control were deemed more trustworthy than systems not providing such information (Verberne, et al., 2012).

The system's information output should be presented in an easy and understandable manner, for instance, through icons. The icons could allow the driver to faster assimilate the content, contrary to messages delivered to the driver through text (Thill, et al., 2014).

In order to have as good feedback as possible it is important to not only use visual feedback but also to use different information outputs aimed for other senses. Two types of information output that have been identified in the literature as favorable to complement visual feedback, are vocal and haptic feedback. In studies made by Toffetti, et al. (2009) it was identified that a vocal interface, consisting of a simple display and vocal messages, shortened reaction time and was considered safer, more understandable, and less distracting by the users than an acustic interface, consisting of an acustic signal in form of a beep and a more informative display. The downside with the vocal interface was that it could be considered annoying by some users (Toffetti, et al., 2009). Another problem with vocal interfaces can be that different people react differently to vocal messages. Accents that differ from one's own can be perceved negatively and create ditrust compared to accents that are similar to one's own, which can increase the level of trust (Waytz, et al., 2014). These findings indicate that perhaps the type of auditory interface should change over time or that the users should be able to customize it.

Force feedback, also called haptic feedback, is a technology that uses vibrations and resistance in control products such as joysticks. Through this type of technology it is possible to receive feedback both from the vibrations as well as through the resistance, i.e. trying to do something but the system pushes back or makes it harder for the user to perform an action. There have been studies on how force feedback could lead to better performance since it allows more fast and exact control over the vehicle as well as making different actions more effortless (Abbink, et al., 2012). Force feedback could be implemented in autonomous systems in vehicles in order give information about what the systems actually are doing as well as hinder the driver from forcing the system to do something that is detrimental, i.e. misusing the systems design intentions. There are also studies that show that haptic feedback can lower the amount of visual feedback needed in order to present information, thus reducing the need for visual input which is beneficial since only visual feedback could be distracting (Abbink, et al., 2012).

This illustrates the importance of using different system information outputs. Providing good information and feedback is especially important in the early stages of the system interaction but will be less relevant the more experience the user gets with the system (Verberne, et al., 2012).

Error Information

Watching an automated system make an error will lower the user's level of trust in the system. The effect of the loss of trust could be reduced by giving an explanation of why the errors occurred and, if it is the case, how they do not affect the overall system performance (Dzindolet, et al., 2003). This could even increase the level of trust (Hoff & Bashir, 2014). The understanding of why errors occur is also especially important when the user is new to the system and still gaining skill and knowledge (Stanton & Young, 2000). The studies made by Dzindolet, et al. (2003) show that the most proper level of trust is reached by not showing the system's obvious errors and at the same time continuously showing the system performance. These authors, however, also discuss that this scenario can be incredibly hard to achieve outside a laboratory.

Uncertainty Information

Studies have shown that drivers provided with uncertainty information have a more proper level of trust in the automated systems (Helldin, et al., 2013) and greater acceptance that the system is imperfect (Beller, et al., 2013). Beller, et al. (2013) also identified other benefits with providing uncertainty information, which are better situation awareness, directing the user's attention to the driving task in safety critical situations, and better knowledge of fallibility, making users better prepared in takeover situations. It is also suggested that the uncertainty information could be presented in a separate view so the user can get the information on demand, making it less intrusive (Davidsson & Alm, 2009).

Regarding how uncertainty information could be presented through interfaces one study points out that only weakening the colors was not enough to represent system degradation but it would also need to be some kind of sound or haptic cue to effectively present the uncertainty information (Helldin, et al., 2013). Another study points out that it is probably good to have a generalized warning, not presenting factors causing the problem, especially in more complex situations (Beller, et al., 2013).

Why & How Information

Information about the vehicle's autonomous system choices is preferable, especially information that is offered in a way that allows the driver to fully understand the intentions of the system. Why & how can be seen as the information regarding the system's upcoming actions, where how information informs how the system will solve a pending task (i.e. braking) and why informs why it will do the task (i.e. upcoming obstacle). The *how* and *why* information have great impacts on the driver's behavior and attitude (Koo, et al., 2014).

It is, however, debatable if both *why* and *how* information are necessary or desirable. On one hand, in a study by Koo, et al. (2014) it was found that only presenting *why* information for the driver is more favorable since it is more prominent and the driver can comprehend the information faster. It was also found that *why* information was most efficient in creating trust. On the other hand was it found that presenting both types (*how/why*) of information is most effective regarding safety, but it could also cause driver anxiety (Koo, et al., 2014). It is discussed by the authors that the anxiety can be because the driver needs to process double messages from the system. It can be hard for the driver to understand which information is most important to process first, the machine status (*how*) or the situational status (*why*), which could cause anxiety in the long run.

Training

One important factor is training or pre-training in order to let the driver understand what the system is capable of in order to not create negative user behavior (Saffarian, et al., 2012) (Toffetti, et al., 2009) such as under- or overreliance which could lead to disuse and misuse (Parasuraman, et al., 2008).

A learning period is needed either through instructions or usage or preferably both (Dzindolet, et al., 2003). Therefore would it be preferable if pre-training could be conducted in such a way such that the user gets to try out the real system in advance in a safe environment without any critical situations. Furthermore, training methods should focus on both system functions as well as the user's different preferences (Merrit & Ilgen, 2008).

But it is not only important to have training before first real use (pre-training). It is also positive to continue this training in order to be able to handle unforeseen situations (Toffetti, et al., 2009). Here it could be beneficial if the system itself could help train the user to widen the user's knowledge about the system's functionality in order to create a more appropriate level of trust.

Common Goals

Establishing common goals with the system could be good in order to align the purpose of the system with the user's. One way to achieve common goals could be for the system to propose certain objectives to the driver, which he or she could accept or decline. If these objectives are mainly positive, a final goal that has been set "together" could also be favorable for the driver (Verberne, et al., 2012) (Lee & See, 2004). These common goals could regard different driving styles such as sportiness or economy driving with less energy consumption as a positive effect (Davidsson & Alm, 2009). In a scenario where goals are shared between the system and the driver it is especially important from a trust perspective that the system lives up to the goals since unachieved common goals could damage the trust for the system even more (Verberne, et al., 2012). To set common goals, share the control with the system and to continuously get adequate and well-presented feedback on how these goals are met could be highly positive for the driver.

Adaptive Automation

By adaptive automation is meant systems that could, by themselves, adapt according to drivers' cognitive and physical preferences (Hancock, et al., 2013) (Helldin, et al., 2013). These adaptive systems that take the human physical and psychological state into consideration could have a great positive impact on human machine interaction (Lee & See, 2004).

Different kinds of adaptive systems could help the driver in many different ways. The adaptive system could primarily decrease the number of accidents through a reduction of overreliance as well as assist the driver in failure response and detection, especially in systems that have different levels of automation as well as manual control (Lee & See, 2004). This will be done through letting the system decide which situations need what types of control, manual or autonomous control (Hoff & Bashir, 2014). But it is also very important to let the system adapt to different kinds of users since people have different types of mental models and therefore different types of trust, tendencies and preferences (Helldin, et al., 2013).

Adaptive interfaces could help lowering the mental workload through only showing relevant information that is needed for the driver (Saffarian, et al., 2012). It is possible to also use different types of design filter features in order to present instantaneous information about workload needed for that specific situation (Saffarian, et al., 2012) which could help the driver to understand what is needed from him or her as well as get a higher understanding of the situation that the system is needed to help manage. Regarding system adaptation it could also be beneficial if the system could learn what the driver likes and wants based on the history of driver choices (Davidsson & Alm, 2009). The autonomous system could learn from the user's driving style for example if he or she usually uses a certain lane at a certain speed and the system could use these preferences and by that after a while only suggest these type of speeds and lanes.

Anthropomorphism (Human-like)

Another continuing step towards a more accepting and positive HMI design is through allowing the autonomous system to be anthropomorphic, as implementing humanistic features in the system (e.g. a name, gender and voice) increases trust (Waytz, et al., 2014). Anthropomorphism affects us positively through being viewed as more competent when it gives the impression of a higher mental state than a machine, through human abilities (Waytz, et al., 2014).

A part of building trust through anthropomorphism is by a two-way communication between the driver and the autonomous system in the vehicle, which leads to a greater understanding as well as a trustworthy platform for the driver to establish trust for the system (Fishmana & Khanna, 1998). This communication could be done through implementing different human abilities, for instance human voice, which could enhance the driver's view of the system's competence (Waytz, et al., 2014). Designing and creating an autonomous HMI system that has human abilities such as human voice, as well as a personality, could according to Waytz, et al. (2014) increase the trust for the system and the autonomous vehicle itself. This is also supported by Hoff & Bashir (2014) who state that a human voice that is "non-interruptive and patient" (p.17) increases the level of trust. It is important to also adapt these voices according to the driver since more dominant users want more direct feedback and submissive drivers may want to hear less direct "commands", as in "do this" and "you may do this" respectively (Lee & See, 2004). Humans tend to sometimes respond to machines as they would have responded to a fellow person (Lee & See, 2004) and that is why it is important to let dominant drivers get dominant feedback from the systems (and submissive drivers get less direct "commands") in order to be more accepting of the system. Other human attributes that are identified to have effects regarding trust are system politeness as well as more anthropomorphic features such as eye shape and its movements as well as chin shape (Hoff & Bashir, 2014). When systems are implemented with anthropomorphic features, they seem to be more competent and thus instill more trust (Waytz, et al., 2014), which also counteracts automation misuse, i.e., not operating the systems in the way it was designed to be used (Hoff & Bashir, 2014).

Customization

It is also important to allow the driver to get information that is tailor-made for him or her through active choices, even after a period of use, as this could enhance the feeling of control (Verberne, et al., 2012). This can be called customization and is different from adaptation since customization is an active choice by the driver and adaptation is when the system itself learns from the driver's preferences. The active choice (customization) could be that the driver does not want to hear certain system information any more, and could therefore turn it off. This is also vital as people are different and want different things, and customization could help to meet these requirements as well as help different users to correctly adjust their trust for the system (Merrit & Ilgen, 2008).

Customizable displays and overall autonomous systems may be personally adjusted to fit the user but different setting adjustments could lead to driver confusion (Saffarian, et al., 2012), which is not preferable. According to Davidsson & Alm (2009), experts consider that it should be possible for the driver to choose the automation level as well as be able to get different levels of guidance in order to facilitate an operator environment that is optimal for the driver through getting the information that is needed for that certain person. Davidsson & Alm (2009) also present information that show that it is relevant for the driver to be able to decide if he or she wants to see system intention or not, which correlates with the customization idea.

Expert & Reputable

It has been found that competence is a very important factor to build trust (Muir & Moray, 1996) and that people trust systems that are portrayed as expert/reputable more frequently, but also that the level of trust declines much faster if the expert/reputable system makes an obvious error (Hoff & Bashir, 2014).

Research has also shown that aesthetics play a large role when it comes to trust in automation. The appearance of the interfaces also highly affects trust even if there is no correlation between the aesthetics and the performance (Lee & See, 2004). This makes it not only important to consider the functionality of the system but also how it looks.

3 Methodology & Implementation

This chapter describes the methods used within the design process as well as how they were implemented for the entire thesis work.

3.1 Design Process

The design process can be characterized by three factors according to Johannesson, et al. (2013), namely, iteration, integration and innovation. These three factors are needed as the design process often presents complex and divergent problems. By complex meaning that the task that is going to be solved often is vague or has contradictory solutions; and divergent, meaning problems that could have an infinite number of solutions per problem. These types of problems need both a creative and analytical approach as well as a holistic view in order to create a structured process leading to an optimal result, hence using the process performer's subjective input in order to choose the "right" solution. A divergent design process should, according to Johannesson, et al. (2013), include different phases such as "formulate the problem, determine the criteria, search for solutions, validate and choose and finally perform" (Johannesson, et al., 2004. p.56).

Johannesson, et al. (2013) serve as an inspirational process template for this thesis, as the aim of this project is complex but above all divergent, presenting different paths to follow (see section 1.2.3 Project Aim). The main work phases, except for planning, used within this project are: define the problem at hand (aim/goal), collect and analyze data, compose data (scenario/framework), conceptualize and select (a highly iterative process), validate and finally present (see figure 7).

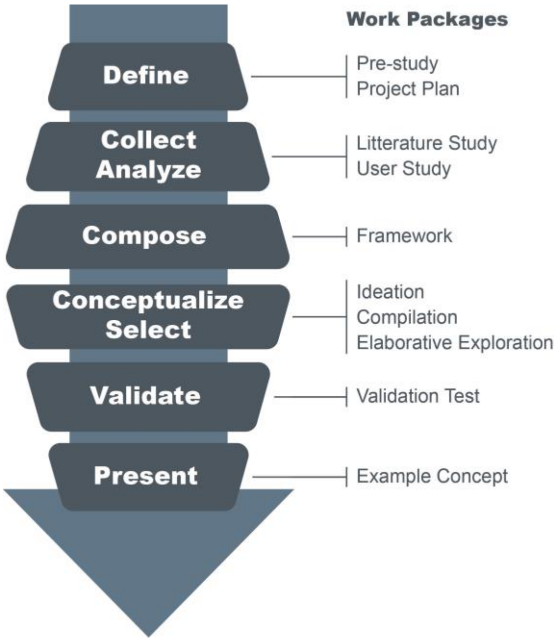


Figure 7 - Main work phases and work packages.

3.2 Define

In order to make a reasonable work agenda within the timeframe, two different planning methods were used: a work breakdown structure (WBS) as well as a Gantt chart. These were then incorporated into a project plan including aim and goals as well as delimitations. This project plan later became the thesis Introduction (see chapter 1 Introduction).

3.2.1 *Pre-study*

A small literature study was performed (here called pre-study) in order to get basic knowledge about the area of trust and automation to be able to define the main issues. The pre-study then formed the background of the initial project plan and thesis proposal.

3.2.2 *Project Plan*

The project plan was done in order to collect all the necessary parts used to be able to plan and structure the oncoming work. The project report contained a background of the thesis work, as well as a planning part using a WBS and a Gantt chart as main planning methods.

3.2.3 *Work Breakdown Structure*

A WBS is a method used to be able to divide the project into different, smaller work packages and tasks and by doing so enabling a better overview of the project as well as helping identify what needs to be done (Eriksson & Lillieskȯld, 2004) (see appendix I). The WBS follows a seven step process: breakdown, identify, work order, time estimation, identify the "critical line" (Eriksson & Lillieskȯld, 2004, p.28), allocating resources, and the last step is to incorporate these work packages into a Gantt-chart.

The WBS was used as a guideline as described by Eriksson & Lillieskȯld (2004) except for not estimating the time for each and every package. This was not done since the time estimation was later done within the Gantt chart instead. A critical line, i.e. a shortest time in which the project can be finished (Eriksson & Lillieskȯld, 2004), was not produced either since this demands that a time is set for each work package. Instead of a critical line, certain milestones were created within the Gantt chart.

3.2.4 *Gantt Chart*

A Gantt chart is an easy to use planning method (see appendix II). It is created through placing work packages on a y-axis and using the x- axis as a time-representing axis (Johannesson, et al., 2004).

The different work packages identified within the earlier described WBS method were implemented into the Gantt chart and given an estimated time of completion. After all work packages had been assigned a time of completion the overall project time could be seen. In order to also know when certain work packages had to be finished, different key dates or milestones were set to have smaller goals to aim for, creating an understanding of how the actual work was following the planned project.

3.3 Collect & Analyze

3.3.1 *Literature Study*

The literature study forms the foundation of the thesis, together with a user study. The literature study was conducted using a grounded theory method using a five-stage process, presented by Wolfswinkel, et al. (2013), which was used in order to obtain a structured research phase. This is because it is believed that rigorous and well-conducted research is important for an optimal end result (Wolfswinkel, et al., 2013). The Grounded Theory, Literature-Review Method (Wolfswinkel, et al., 2013) is a technique that is relevant in order to be efficient within the research phase. The five steps of the grounded theory are: define, search, select, analyze and present.

Define

The first step is to define the relevant area of interest as well as defining which search terms that should be used. The main areas were trust and human-machine interaction. These main areas also had subareas that were researched:

- Trust
  - Interpersonal trust
  - Human-machine trust
  - Trust in autonomous vehicles
- Human-Machine interaction
  - Automation
  - Autonomous vehicles

Search

The next step was to start to search for different relevant sources based on the chosen search terms.

Select

When several articles or similar have been found, they should be read in order to see if there are any unnecessary or redundant articles. The articles could be analyzed by first reading the titles and abstracts in order to see if the article is relevant and then continue on to look at the citations to see if there is any possibility to find new and relevant information sources. This step is highly iterative, since it could entail new articles through the citations and could go on until all relevant articles within an area of interest have been looked at.

Analyze

When analyzing the chosen articles it is favorable to highlight key words and sentences that could be extracted from the articles. These extracted parts are called excerpts (see figure 8) and allow the researcher to group similar information and by doing so start to form a mental image of what the information is stating, creating a holistic perspective over the information gathered.

Figure 8 - Structured Excerpts

Present

The last step in Grounded Theory by Wolfswinkel, et al. (2013) lets the researcher structure (see figure 8) the information gathered from the earlier steps to form the final article.

The literature study was conducted within the areas of human-machine interaction, trust, both between human and human and later also between human and machine and finally, autonomous driving vehicles and their interrelated autonomous driving systems.

3.3.2 *User Study*

The user study was conducted using a NHTSA level 2 autonomous vehicle system found in a Mercedes Benz E-class. The results are first and foremost only used as a guideline to be able to pinpoint where different key factors affecting trust can be best suited in an autonomous, level 3 NHTSA vehicle. Since there were not any level 3 vehicles available for the study, the second best had to be chosen, namely the level 2 NHTSA vehicle from Mercedes. The problematic issues regarding level 2 and level 3 systems should have similarities regarding e.g. control transitions and therefore a level 2 vehicle was deemed acceptable. The system used in the vehicle is called Distronic Plus and also had steering assist. This system can be seen as an adaptive cruise control that both brakes for vehicles lying ahead as well as follows them (distance-keeping). The system can also to an extent help the driver with steering via lateral lane guidance, i.e. keeping within the lane. (Mercedes-Benz, 2015).

Nine people between the ages of 23 to 55, including four females and five males participated, each having a driver's license between 5 and 37 years. The study participants had all a technical background and were all affiliated with Volvo Vehicle Corporation, mostly through master thesis work but also through employment. It should be noted that the participants were first-time users of the level 2 system. The user study also used a semi-structured interview technique, constructed in a manner where the participants had to operate the vehicle on a common road with low-density traffic at the same time they were asked certain questions. These questions covered the whole test drive but focused mainly on certain critical touch points (called events later on in the thesis), e.g. manually driving the vehicle, Distronic Plus and steering assist system activation/deactivation and usage, as well as their thoughts of the system after the test drive. The questions were constructed in a manner to receive as much feedback as possible regarding the participants thoughts and feelings about the level 2 system in terms of trust and usability as well as what is likely to be needed during the usage, and by this trying to gain an understanding of what would be needed in order for the driver to trust and use a level 3 system correctly (see appendix III). For results regarding the user study see section 4 User Study.

3.4 Compose Data

3.4.1 *Framework*

Introduction

From the literature study and the user study, information was extracted regarding trust factors as well as where these could best be implemented in an interaction with an autonomous driving vehicle. Therefore a framework was created (see figure 9). The framework could be seen as a scenario composed of different events (see section 5 Framework for results).

How does it work?

The main idea with the framework, which also is the thesis' main result, is to have easy and accessible guidelines that designers can use when constructing a human-machine interaction system that is used as a bridge between the user and the AD-system. The scenario is deliberately loose regarding design examples, since it should allow the designer to use it as guide as there are several design solutions to one problem.

When specifically speaking about an autonomous driving system, it will be referred to as an AD system or autonomous driving systems. Other systems of the vehicle will only be referred to as system(s). The root cause of also including other systems into the scenario and being required to discuss them is because of that fact that the trust for a machine is reflective of all its parts not only explicitly the AD system. Aesthetics, overall impressions as well as other systems of the vehicle will have an impact of the driver's view of the AD system. The other systems also need to be considered since they could be interrelated to the AD system through different functions (Hoff & Bashir, 2014) (Lee & See, 2004).

Phases & Events

The scenario in the framework is constructed to contain all identified types of events and is intended to be a holistic view of how an interaction with an AD vehicle will presumably look. In most cases the scenario should be followed but could vary from user case to user case, since perhaps not every potential customer or driver goes through all events.

The scenario is constructed of three main phases: the pre-use, learning and performance phase respectively. The pre-use phase treats what happens before the first physical interaction with the vehicle and the AD system. During the pre-use phase two events take place, namely, implicit and explicit information gathering.

During the second main phase, the learning phase, the user gets the first physical interaction with the vehicle and the autonomous driving system. This phase is not isolated to only first time usage, rather until the driver has learned how the AD system works, which could take different amounts of time and effort depending on the individual. In the learning phase eight different events are identified: enter vehicle, activate vehicle, manual mode, control transition one, autonomous mode, control transition two, manual mode two and exit vehicle.

The third and last of the main phases is the performance phase. This phase is based on three different events: continuous usage, change of context and incident.

Change of context could also occur in the learning phase but does only need to be considered after the learning phase or rather after the continuous usage event. This since change of context basically only explains that if the driver has learned the system and an unfamiliar situation occurs, it could be good to give more information again, hence taking the driver back to a small learning phase. The same goes for the incident event, which also could occur during the learning phase as well as during the performance phase. The trust affecting factors will remain the same regardless of which of the two phases it occurs within.

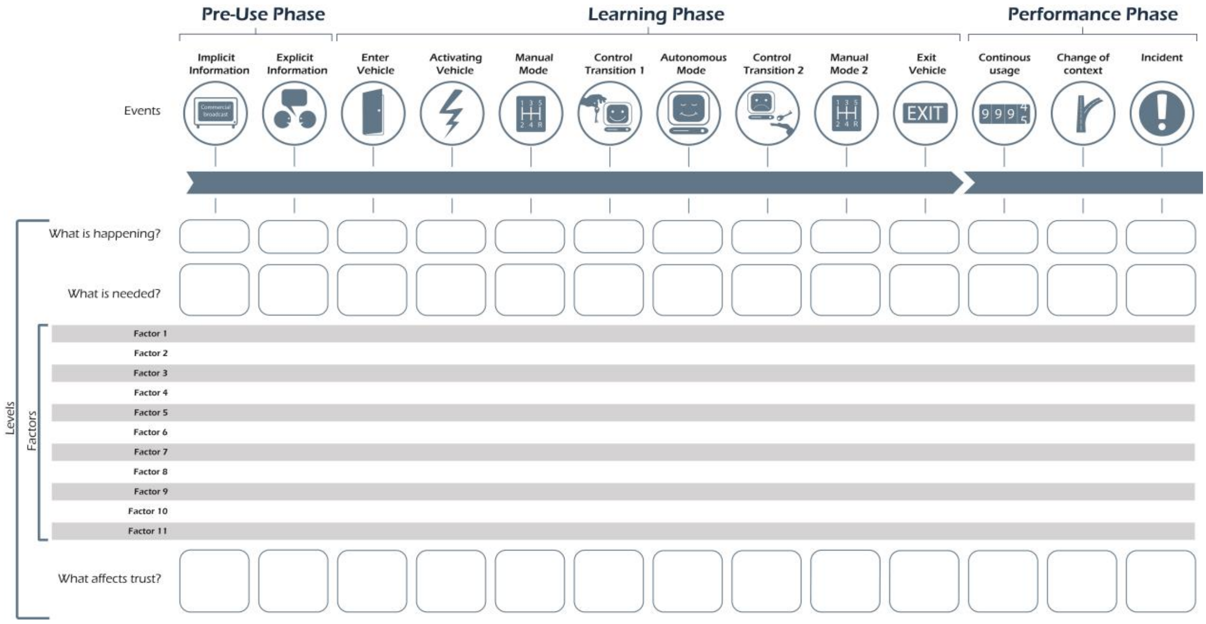


Figure 9 - Framework created as a scenario.

Levels & Factors

In addition to the three main phases and the events, there are three different levels explaining what is happening, what is needed as well as what affects trust (see figure 9). These levels are connected to each and every event in the scenario. The scenario is constructed this way in order to easily understand what tasks are performed as well as what these task needs in order to be able to be properly conducted. The third and most important level is the trust level, i.e. which trust factors could be used in order to create an appropriate level of trust during each event. This level uses the identified trust factors and explains them further within a context.

3.5 Conceptualize & Select

3.5.1 *Ideation*

Two different ideation sessions were conducted, one with students from Chalmers University of Technology (CTH) and one with experts from Volvo Vehicle Corporation (VCC).

The six students that participated from CTH each had engineering backgrounds with a product development and/or industrial design orientation for their master's studies.

The workshop with experts was conducted at Volvo Vehicle Corporation. The four experts were employees of VCC or had consultant employments with VCC. Their expertise ranged from engineers to scientists within in areas such as human-machine interaction to human factors regarding autonomous vehicles.

The ideations workshops were conducted in order to create a lot of ideas regarding how you could create an appropriate trust level for future level 3 autonomous vehicles through an HMI system. The ideation sessions used different methods (discussed further below), since a systematical approach when generating ideas is key in order to keep it structured, as well as achieve an as good result as possible (Michanek, et al., 2007).

Workshop Students

The ideation with the students used two different ideation methods and a more loose discussion. The two methods used were Brainpool and 6-3-5, both brain writing methods, i.e. ideating individually in silence which encourages participation by letting the participants ideate on equal terms, in comparison to brainstorming which is an open discussion that could lead to certain participants being more active than others (Heslin, 2009).

In the workshop, a first, short ideation task was done before introducing the oncoming agenda in order to let the participants get the right mindset before explaining the aim of the workshop in more detail, keeping the participants as open-minded as possible. After the first ideation method was completed, there was an introduction of the theme as well as a presentation on how the remaining part of the workshop would be executed. After the second ideation method, the participants got a short intermission, which was immediately followed by a 60-minute discussion.

The first, short ideation task followed the Brainpool method (see figure 10) that allows the participants to generate ideas based on one or more questions or topics through writing or sketching their ideas on a piece of paper. Then these ideas are put in the middle of the table allowing others to see them and get inspired by them in order to further develop or get totally new ideas from them (Michanek, et al., 2007). The question that the workshop participants had to focus on was "How do you create trust?" and they had only five minutes to generate ideas.

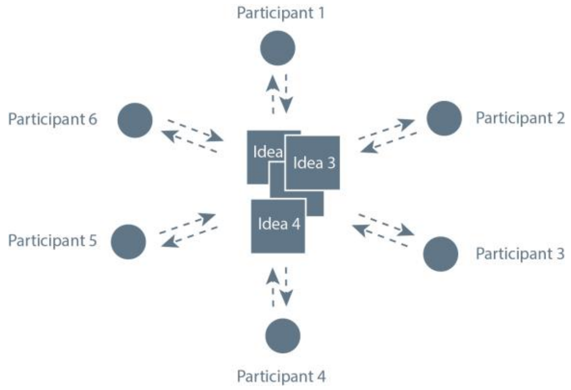


Figure 10 - Ideation method Brainpool

After the introduction a second method was used, namely 6-3-5, which also builds ideas upon the other participant's ideas but in an even more structured manner. The name for this method comes from that it uses six participants who, during five minutes each, generate three different concepts on specific questions and/or topics presented to them. All participants start with their own clean slate of paper with three columns and six rows. After one participant has come up with three different ideas sketched or written in all three columns (see figure 11), he or she passes the paper to the participant that sits right beside him or her.

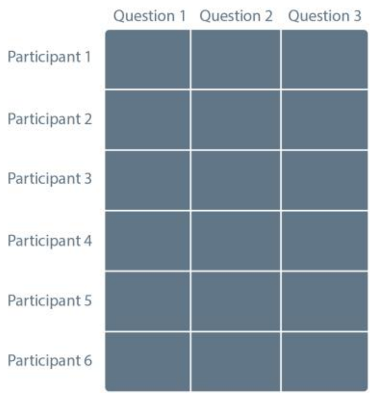


Figure 11 - Ideation Method 6-3-5

The direction of the paper hand-over, clockwise or counter-clockwise, should be agreed upon before executing the method. The person that gets three concepts or ideas from the previous idea generator can choose to further develop these, or use them for inspiration, or just ignore them and continue with his or her own ideas (Michanek, et al., 2007). When all rows are finished, i.e. fully filled, the session is finished, taking around 30 minutes.

The method was used with a small alteration. Since there were six different questions, three of the participants got to start with three questions and the remaining participants started with three other questions. When they started to pass these papers forward, all of the participants got to ideate on all six questions. The questions used for the 6-3-5 ideation were:

- *How can you increase SA?*
- *How can you minimize mode confusion?*
- *How can you make an HMI system more user-friendly through an interface?*
- *How can you make an HMI system that fits everyone?*
- *How would it be possible to increase the “feeling” of teamwork with a self-driving autonomous system?*
- *Which information would you like to receive from an autonomous vehicle system?*

Since these questions were relatively precise and presented very subject-specific words, these had to be explained beforehand in order to allow the participants to have an understanding of what the words and questions really meant.

After the 6-3-5 ideation method, the participants got a short intermission, which was immediately followed by a 60-minute open discussion. The open discussion focused on the topics of Mental Model, Workload, System Transparency, Anthropomorphism and Two-way Communication. There were specific questions asked during each specific topic, which can be seen in the appendix IV.

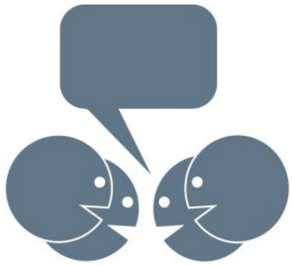


Figure 12 - Open Discussion

The open discussion (see figure 12) was used in order to let the participants to ventilate ideas they had about the topics. They were also allowed to go further in their thoughts and connect these topics to other similar issues to an extent, which could lead to new and interesting approaches. The discussion results were continuously written down, focusing on new innovative approaches.

Workshop Experts

The workshop held with the experts was quite similar even though it had certain differences in comparison to the workshop held with the students (compare appendix IV and appendix V). In terms of the methods, first the same Brainpool method was used, followed by the standard 6-3-5 but with only four participants, from here on out called 4-3-5 ideation method. Then, a third method was added, based on a highly modified 6-3-5 method, using only two columns and four rows (from now on be called the 2-2-3 group method).

The questions were also a bit different focusing more on certain aspects of interest that were not as prominently featured in the literature study. Thus it was deemed relevant to discuss these aspects with people that work with these types of topics on a daily basis. Since there were only four participants in this workshop, the methods had to be modified to better match the ideation task.

The Brainpool method (see figure 13) that started the workshop used the same question as in the workshop for the students namely, “How do you create trust?” The time limit was also the same, five minutes. Since the participants in this session had more knowledge about the topics, the introduction after the first method became unnecessary.

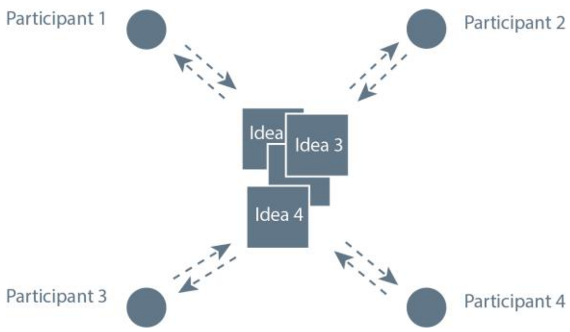


Figure 13 - Ideation method Brainpool

The method called 4-3-5 followed, but since there were only four participants, the number of rows was reduced to four (see figure 14). The time limit was also reduced accordingly, to 20 minutes. The questions were limited to three, namely:

- *How can you increase SA?*
- *How can you make an HMI system to fit everyone?*
- *Which information is needed for the driver in an autonomous vehicle system?*

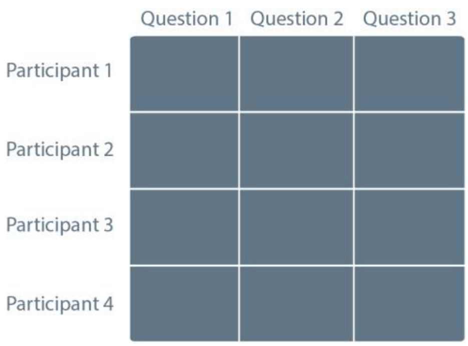


Figure 14 - Ideation method 4-3-5 (6-3-5)

After the 4-3-5 method, the participants took a ten-minute break which was followed by the 2-2-3 group method (see figure 15). In this exercise, three different topics with associated questions were addressed, namely:

- *Mental models*
    - *How should you present the system in order to make it look like an expert?*
    - *How can you create a correct mental model before first use?*

- *Workload*
    - How could you minimize stress but still keep the driver alert?
    - How would you like to use your different senses (haptic, visual and auditory) in order to receive system information?

- *Anthropomorphism*
    - How would an anthropomorphic system look?
    - What kind of features would make you trust another person?

In this ideation method, the four participants were divided into two groups of two. They were then handed one sheet of paper each, similar to the paper in 6-3-5 but instead with two columns and four rows.

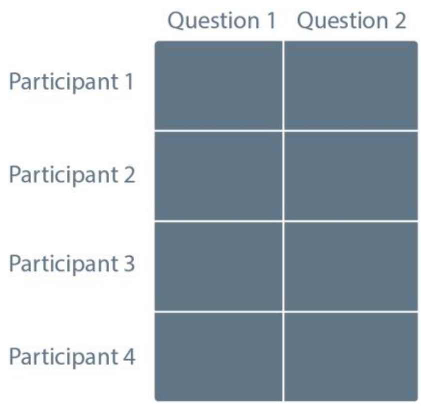


Figure 15 - 2-2-3 (6-3-5) Group method

Every column had a question connected to a topic that every participant should ideate around in his or her own group. They got three minutes to ideate two concepts connected to each of the two questions, then they had to pass the sheet of paper with the two concepts to the other group member (see step 1 in figure 16). After 12 minutes the sheet of paper was full and both groups got the chance to sit down, still in their own group, and for ten minutes conceptualize (through discussion) an idea based on the smaller concepts ideated on the two sheets of paper (see step 2 in figure 16).

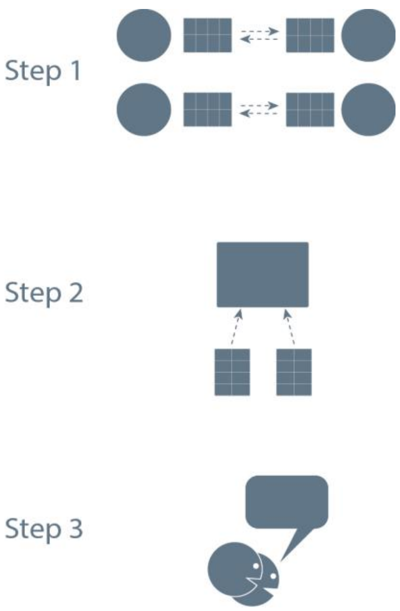


Figure 16 - 2-2-3 Group Method Process

The two groups then got ten minutes to present and discuss the concept that they had developed for the other group as well as for the mediators (see step 3 in figure 16). This process was repeated three times treating three different topics with associated questions (see above). The full procedure can be found in appendix V.

3.5.2 *Compilation*

Affinity Diagram

During the compilation phase an affinity diagram, also called KJ-method, was used in order to structure the information gathered from the ideation phase. An affinity diagram is favorable for clustering a lot of information and to create associations between groups or themes (Bergman & Klefsjö, 2010). The affinity diagram explained below is based on Bergman & Klefsjö (2010) but was executed with several differences.

The method starts with defining the focus, i.e. posing questions. The questions should then be answered by the participants (in this case the project owners, or the authors of this thesis) through silently writing down on post-it notes or similar what, according to them, are solutions to the questions.

The post-it notes should then be put on a wall for everyone to see in no particular structure. When all post-its are on the wall it is time to go through the different answers clustering them together, something that is done together within the group If there are any redundancies they can be removed. The clustering process, placing different post-it notes with different answers in cluster groups based on the meaning through association, should be done in silence. Conflicts are bound to happen since the participants within the group are probably going to have different opinions regarding the meaning of a certain answer and their connection to a certain cluster group. The conflict will however according to Bergman & Klefsjö (2010) decrease as the clustering process continues.

When all post-its with respective answers are put into clusters creating several groups the work is complete. When this has been done it is time to create a cluster group name that should in best way possible convey the core meaning of the entire group. After this it is also possible to draw arrows between different cluster groups, hence creating connections between them. It is also possible to structure all the groups in a hierarical fashion by giving them different numbers and by doing so giving certain groups a higher importance level than others.

The affinty diagram method used during this compilation process was applied in a different way than the above explained process. Since all the ideas, or solutions as they should have been called according to Bergman & Klefsjö (2010), already existed through the earlier ideation, the ideas from the ideation workshops were structured using this method as a template. Since there were a lot of ideas, the method had to be repeated two times: initial and final analysis.

Initial Analysis

During the initial affinity diagram analysis, all the ideas from the ideation sessions were divided into different clusters (see figure 17) based on their similarities (see appendix VI).

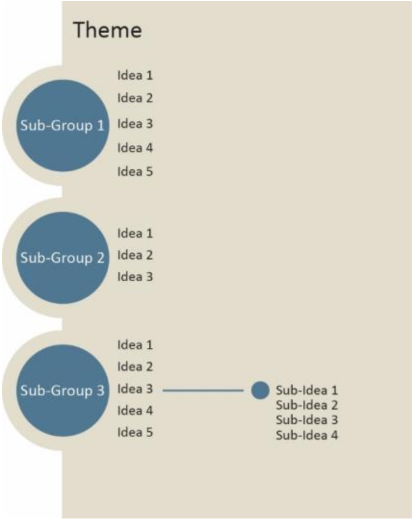


Figure 17 - Clusters in Affinity Diagram.

In order to be able to even categorize all the ideas into the initial analysis, a lot of discussion between the project owners took place. These discussions when categorizing the ideas into themes were more about subjective connections than logical (Bergman & Klefsjö, 2010), hence where did they fit the best according to the project owners.

Final Analysis

Since the different themes in the initial analysis still were too many to grasp, a second and final analysis had to be done in order further decrease the number of themes and their connected ideas. This analysis followed the same pattern as the first analysis.

After the second and final analysis, a new phase took place, which was more of an elaborative exploration of possible combinations of the ten different themes and their connected ideas in order to make them more illustrative and comprehensive.

3.5.3 *Elaborative Exploration*

Initial Phase

During the elaborative exploration, the first step was to illustrate the ten themes (resulting from the affinity diagram exercise) and their ideas through sketching combinations. Each of these was then put into a matrix with a determined scale in order to organize and more easily visualize the ideas. The scale (see figure 18) was used to separate similar ideas by categorizing them for example by how anthropomorphic the ideas were.

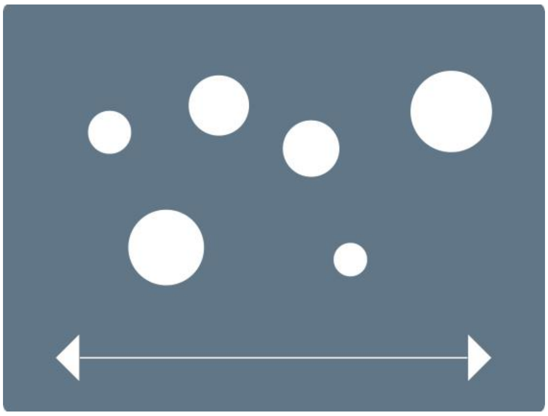


Figure 18 – Visualization over the matrices with scale.

Evaluation 1

After the ideas had been weighted in reference to each other, ideas were selected through a discussion between the project owners with the framework in mind. Since all of the ideas were thought to more or less solve the trust issue, it was more about how well they could work together. The different ideas were then put into one of three different main HMI concepts.

Final Phase

The three different main concepts, from now on called concepts one, two and three, were further illustrated by placing and connecting them (represented as balloons in figure 19) illustratively to a vehicle's interior together with small explanations of what and when they did something that could affect trust in a relevant way. The same procedure was done for all three concepts. The results regarding concept one, two and three can be found under section 6.3 Main HMI Concepts.

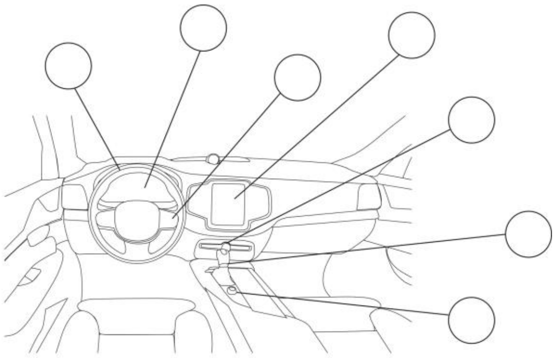


Figure 19 – Visualization of how the elaborative exploration connected the concepts to the interior, creating three main concepts.

3.6 Validate

3.6.1 *Introduction*

A validation test was conducted in order to validate the factor 'feedback' during autonomous mode (AD mode). This factor and event were chosen because the project owners considered it to be interesting to test a factor within the event 'AD mode' or an event connected to it. But since there were no autonomous driving vehicles to be used, control transitions were not possible; however, a fake AD mode could still be obtained within an ordinary vehicle. The factor 'feedback' was chosen since there were ideas that were relatively evolved and could be implemented into the validation test fairly easily. These ideas were drawn from the three different main concepts (one, two and three) found in section 6.3 Main HMI Concepts. These ideas focused on object recognition and had different levels of transparency feedback. By using these three object recognition systems, from now on called OR-concepts, it became possible to see which transparency level is favorable as well as validating the framework at the same time.

In the literature study (see Transparency section under 2.2.2 Issues), it was identified that a transparent system will give the user an increased feeling of control since the user has a greater ability to predict the system's behavior (Verberne, et al., 2012). In turn, this gives a sense of knowing what the system is "thinking". Since much of trust is based on observations of behavior, it is a crucial factor that the behavior is made observable (Muir, 1987).

This led the project owners to formulate the hypothesis that by showing transparency feedback through object recognition, the driver will get an understanding about the system's process leading to an increased level of trust. It is also believed that the feedback will give a higher level of situation awareness in takeover situations since the driver will get an increased understanding of the surroundings.

3.6.2 *Research Method*

The method used in the validation test was the repeated measure design approach where all participants get to try each condition of the evaluation. This approach needs fewer participants and is efficient since each participant will test all the different conditions. It is also sensitive since it can detect even small variable differences between the conditions (Shaughnessy, et al., 2000). The downside with this method is also what makes it good, the fact that the participants get to try the scenario several times with only small adjustments. This could lead to a phenomenon called the practice effect that may lead to improvement in performance or that people become bored and tired over time (Shaughnessy, et al., 2000). The order in which the conditions are tried will therefore be factor affecting the result of the evaluation, although it can be counteracted by changing the order in which the conditions or the concepts occur. One way to do this is by using a Latin Square Design where the orders are pre-selected with the help of a matrix where all concepts occur in every start order one time. The matrix used in the validation test can be seen in table 1, where A = OR-concept 1, B = OR-concept 2, C = OR-concept 3, D = no concept.

Table 1 - Latin Square Design

A	B	D	C
B	C	A	D
C	D	B	A
D	A	C	B

3.6.3 *Participants*

There were totally 8 people (7 male and 1 female) with an age ranging between 19 and 28 participating in the validation test. None of the participants had a technical background and were randomly assigned in which order they tested the concepts.

3.6.4 *Procedure*

The validation test was conducted in a right-hand driven car with a curtain attached between the driver and passenger seat so that the driver could not be seen by the passenger. Each of the three OR-concepts were designed as an animation in Adobe Flash Professional cc 2014 that were displayed on a tablet that was placed on the dashboard on the passenger side.

The OR-concepts were controlled by one of the test leaders, sitting in the backseat, with the help of a computer, and the other test leader drove the car. The setup can be seen in figures 20 and 21.

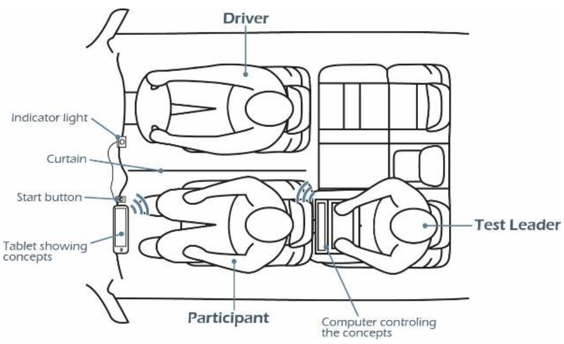


Figure 20 - Visualization of validation test setup.

Figure 21 - Picture of the interior setup used in the validation test.

The participants were informed about the purpose of the validation test and told that the car used in the test was autonomous and programmed to drive a predefined route. They were also informed that there was an object recognition software connected to the car's sensor system and that the information was presented as different object recognition concepts shown on the tablet.

After the introduction the participants ran the course four times with a few minutes break between each lap. Each participant tested all three different OR-concepts in separate laps as well as one control lap without any object feedback at all. The order in which the concepts and control lap where tested was randomized using the Latin square matrix. After each lap the participants got to fill in a 'trust in automation' questionnaire developed by Helldin, et al. (2013) which in turn is based on the empirically determined questionnaire by Jian, et al. (1998). The questionnaire contains 7 questions that are answered using a 7-point scale where 1 is "I totally disagree" and 7 is "I totally agree" (see appendix VII). After all four laps were completed, the participants were also interviewed about their feelings about the test and the different concepts.

The course was approximately 600 meters long and was located on an empty parking lot. Each lap consisted of four object encounters: overtaking a moving car, following a moving car, avoiding a cone to the left and avoiding a cone to the right. A visual representation of the course and object encounters can be seen in figure 22.

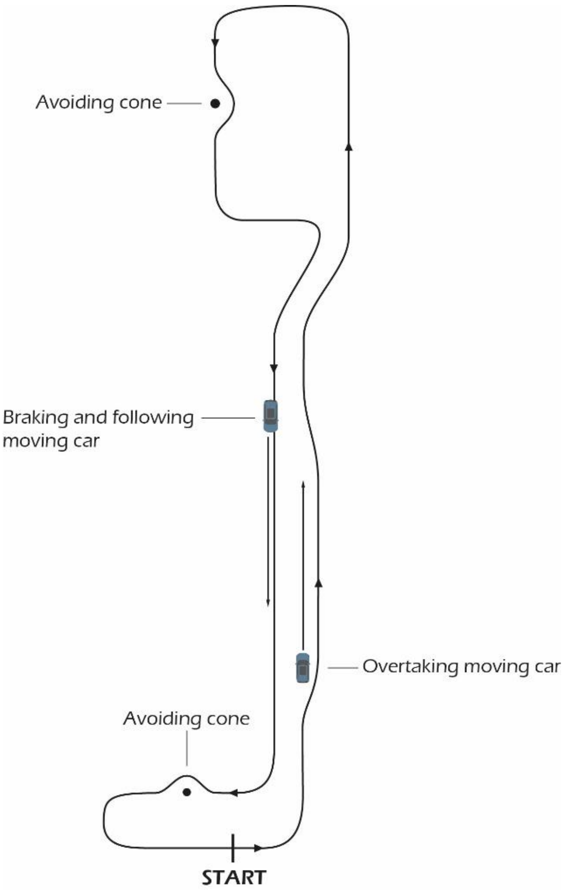


Figure 22 - An overview of the course and the object encounters with an example of outcomes.

The encounters were positioned in the same place of the course for all participants so each lap would be almost exactly the same. One variation was though that the outcome of each encounter were randomly decided (e.g. if the car will avoid the cone to the left or right) in order to have some small variations between each lap to give the participants a small degree of uncertainty of what would happen. The data collected from the questionnaire were afterwards statistically analyzed.

3.7 Present

The presenting phase uses illustrative pictures with the goal to convey a final example HMI concept. The pictures are actually stills from animations showing how the example concept works, as animations are not possible in this thesis. The example concept is based on the three different main concepts created during the development phase. The example concept is only to be used in order to get an understanding of what types of results the framework can produce. The only part that is validated within the example concept is the OR-concept, affecting the factor 'feedback' during the event 'AD mode', for which results can be found in chapter 7 Validation Test.

The example concept only treats the learning and performance phases, i.e. the interactions between the driver and the autonomous driving vehicle and its systems. Even if the example concept only focuses on the two phases that treat the direct interaction with the AD systems and the autonomous vehicle, the pre-use phase is just as important but only the physical interaction within the car was regarded in this thesis.

4 User Study

This chapter presents the result from the user study. The user study was conducted as a part of the research in order to further identify or validate important trust factors and correctly place them in a scenario-based framework. A level 2 vehicle was used only as a reference to see what the designer should focus on in a level 3 vehicle regarding the trust aspect.

After the different solutions had been drawn from the research regarding HMI factors indirectly affecting trust and more exact trust factors, a user study was conducted in order to guide the placement of the key factors into the driving scenario-based framework, in which the different driving events were to be connected to the key factors affecting trust. In other words, the user study helped identify where different trust factors and their solutions could be most effectively applied in the quest for constructing an appropriate level of trust for the level 3 autonomous vehicle´s HMI system. For interview guide see appendix III.

4.1 Introduction

The interview that was conducted during the user study was done so during a test drive within a level 2 vehicle, namely a Mercedes-Benz with Distronic + and steering assist. The interview results are presented both in quotation form throughout the text as well as in a more loosely compiled form (see figure 23).

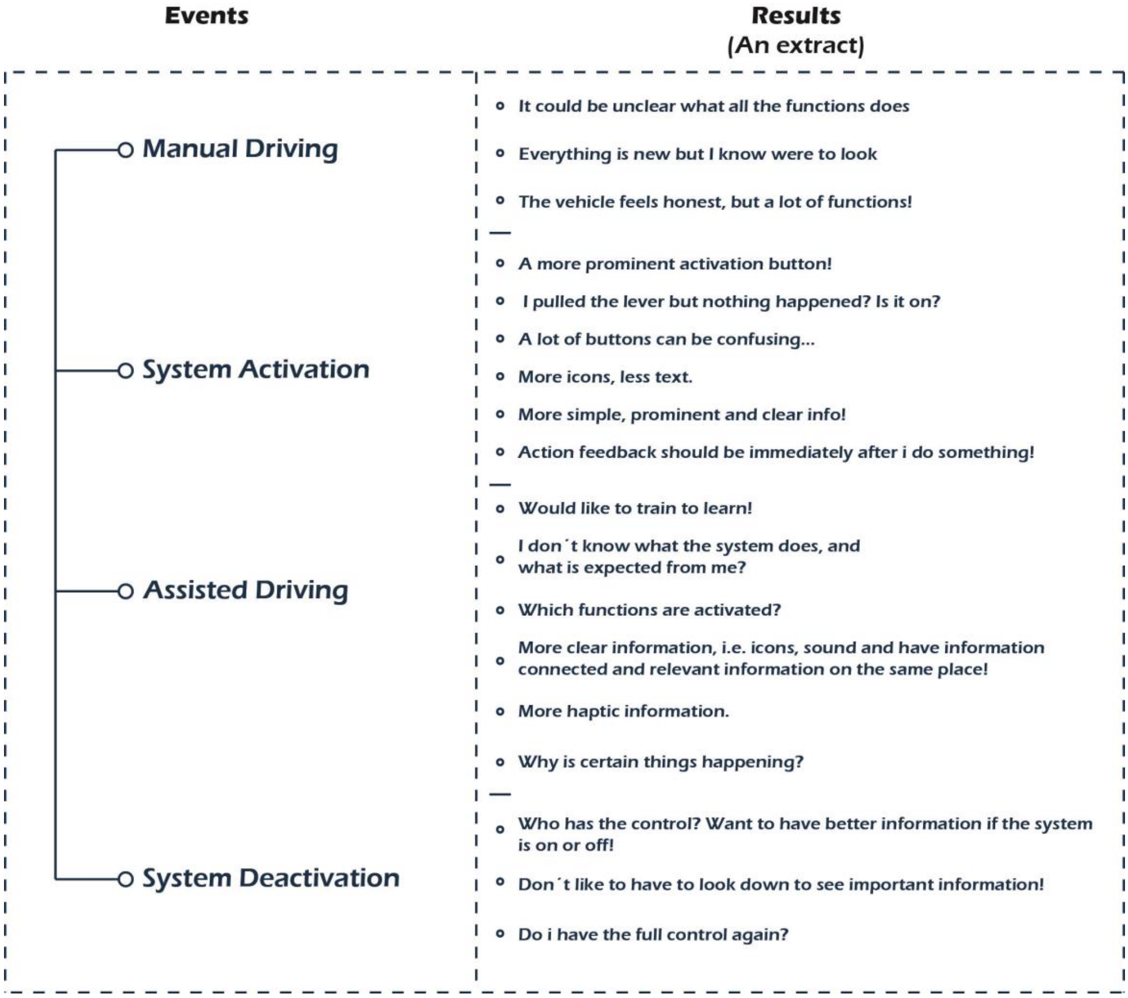


Figure 23 - Result presented in a compiled form.

4.2 Manual Driving

During manually operating the vehicle, the test participants did not really see a problem with the standard systems affecting the drivability of the vehicle. The participants felt quite relaxed overall, though one person added that it was "a lot of buttons, which could be confusing". On the question of how familiar they felt with the system, one person said that "everything is new but I know where to look" meaning that the vehicle and the visual design of the systems were new for her but since she had driven other vehicles before she knew where she needed to look to get certain feedback regarding manually operating the vehicle.

4.3 Assisted Driving Activation

The assisted system activation can be seen as a transition between manually driving the vehicle and getting assistance from the system in the form of braking, controlling speed and steering (lane keeping). During this event, several questions were asked, for instance; what did you feel when activating the assisted driving system? During the assisted driving activation most of the participants had problems realizing when the system was fully operational. One person responded: "I pulled the lever but nothing happened!" In actuality the Distronic Plus system had turned on but almost all participants missed the icon indicating that the steering assist was activated. (The icon was a small steering wheel placed next to the speedometer and the tachometer which turned green when the steering assist was fully engaged and grey when it was not). So when system was active but the steering assist was not engaged because the system had not found any lines or a vehicle to follow, the user became puzzled about if the system was on or off.

When asked about what information the participants would have liked to have more of, one person said "more icons and less text" which most of the participants agreed with. There actually were icons but as said before, the steering wheel icon was mostly overlooked causing many to fail to recognize it. It was not until the information provided to them by the interviewers that they noticed that icon. There was also information mediated through text but, as a participant commented, it must be clearer and more understandable in that way that it explains more exact what functions are active, and preferably having all information on the assisted driving in one place

4.4 Assisted Driving

During the assisted driving mode, which should not be confused with fully autonomous mode, the participants felt that it was hard to understand if the system was on or off (the same issue as in the assisted system activation event), what the system will do, and what the driver needs to do and have control over. For the question on what would give more trust for the system, several participants said that they would have wanted more information about when the system is fully activated, what it does, and in general more information about the system. They also said that they would have wanted to learn the system before first usage in order to familiarize themselves with the functions.

When asked questions regarding how they would like to have the information presented, several participants said that they would like the visual and audio information to be more clear – as one participant noticed when a beep sound occurred, "the beep was unclear, maybe use some other kind of information instead". Participants also wanted more haptic feedback i.e. vibrations regarding upcoming events and warnings etc.

4.5 Assisted Driving Deactivation

The system deactivation faced almost the same problems as did system activation, especially regarding the earlier mentioned issues such as knowing when the driver is in full control again and which functions in the system are still active (only Distronic Plus or both Distronic Plus and steering assist).

4.6 General Thoughts

The overall issues regarding the HMI system's usability and trustworthiness are the lack of prominent and clear information (especially if the system is on or off), what it will assist you with as well as why it does certain actions. According to the study participants, other issues regarding the HMI was the placement of different functions e.g. the activation lever which was located behind the steering wheel and by that totally hidden from the driver. The people that tested the system had a preconception of finding the Distronic Plus and steering assist activation function somewhere on the steering wheel, which it was, but behind the steering wheel on the steering column. The conclusion that can be drawn is that the activation function could be on the steering wheel or somewhere else entirely as long as it is very prominent.

Another issue some had was that they had to look down to get information regarding the assisted driving, which was not very popular since they wanted to continue to be able to have their eyes on the road.

5 Framework

This chapter is based on the results from both the literature study and the user study. The result is a framework developed as a scenario showing where the different trust-affecting factors are best applied within the interaction between the driver and the level 3 autonomous driving system.

5.1 Introduction

The scenario starts even before the first physical interaction between the driver and the HMI system in the autonomous vehicle, this is because trust starts from the first communicative explanation of the system that creates expectations that could be right or wrong. A graph called Optimal Life Cycle of Trust (OLCoT) (see figure 24) is presented explaining how the project owners interpret how a human-machine interaction should optimally alter the level of trust. Something to be noted is that the graph is only a very simplified illustration of the trust formation process.

The OLCoT graph uses the exact same main phases as the framework, namely Pre-use phase, Learning Phase and Performance Phase, and shows how trust changes from before the interaction to when the driver fully understands the system. The graph is created to explain how trust is formed by usage, and is an attempt to simplify how the trust formation process works. In the pre-use phase, the graph shows that the level of trust should not be as high as possible but should be on a level so that all users operate the system in a correct way from the beginning, not misusing or disusing the AD system.

In the learning phase, that starts from first usage and continues until the user has learned the AD system, it is up to AD system to form the level of trust in reference to the system's actual competence, i.e. towards an "optimal" level of trust.

After the user has learned how the system works, trust is mostly based on the performance of the system and the level of trust is stabile unless something happens. Two different events have been identified where

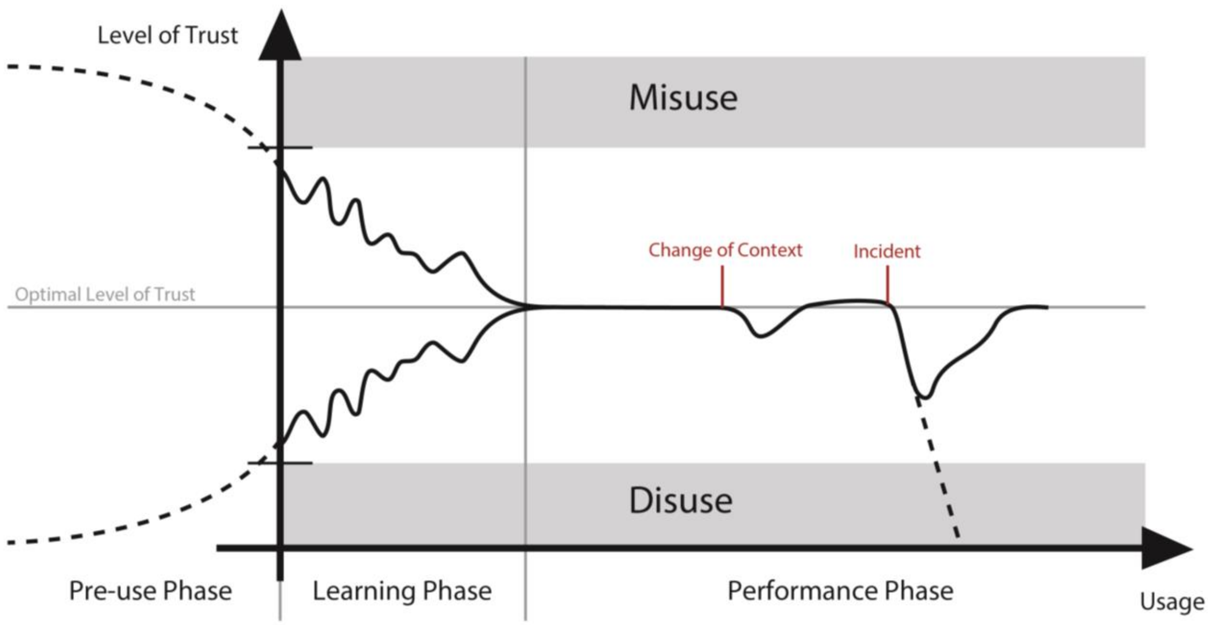


Figure 24 – OLCoT graph

the system may need to compensate for a fluctuation in the trust level. These two events will be further explained below in section 5.4.2 Change of Context and section 5.4.3 Incident.

With the graph in mind, the framework explains the different events taking place, i.e. a scenario, from before the first physical interaction with the AD system and the vehicle to when the driver has learned how to operate the AD system.

The presented framework (see figure 25) is intended to be used by designers creating new HMI systems in autonomous vehicles, especially level 3 AD systems. It could also be used in order to validate existing systems to see if they create the amount of trust sought after.

The next sections explain the framework phases, events, and matching trust factors.

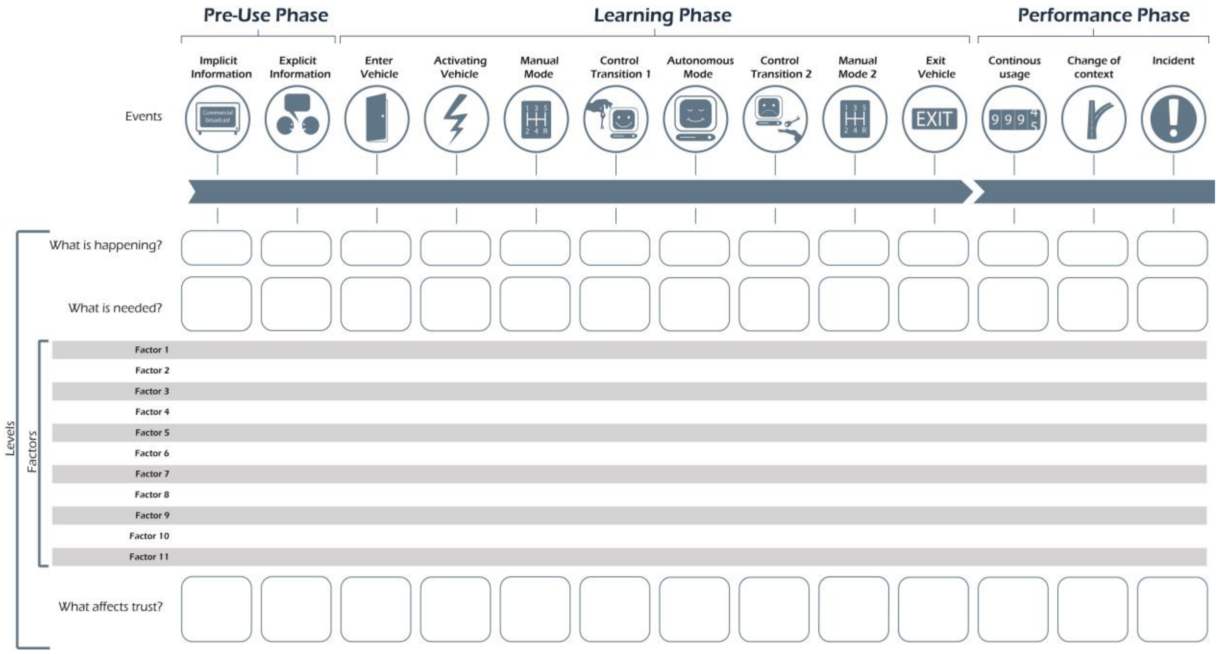


Figure 25 - Framework

5.2 Pre-Use Phase

5.2.1 *Implicit Information*

What is happening?

Implicit information is presented, meaning information that is given to a passive recipient, i.e. a person not actively searching for the information him- or herself. In this case, a potential future user is presented with information about the AD system. This could be done through commercials or through word-of-mouth communication.

What is needed?

Since the recipient is in a passive state of information exchange, he or she therefore only needs to make a choice regarding how to proceed with the acquired information, choosing between seeking out more information or not.

What affects trust?

It is important how the company's or designers' intentions are shown, for instance through commercials communicating the purpose of the AD system to its possible future users. It is highly important to only show a purpose that is coherent with the actual performance of the system or else the trust for the system can be severely diminished during future usage. An example could be if the AD system is portrayed through commercials as being able to drive on its own in high-traffic urban areas, but in reality could only be used on highways with lower traffic density and the customer realizes this, the outcome could be a negative impact on the user's trust for the AD system.

The AD system should also be presented as a competent and reputable agent, hence focusing on its strengths as a system, i.e. letting the possible user to know how good this system has performed, for instance through different AD system tests in which the system has passed with flying colors (see section Expert/Reputable in 2.2.3 Factors).

It is also very important to let the possible user to know that his or her goals are similar with the goals of the AD system, hence allowing the customer to feel that the systems intention(s) correlates to his or her own view of what is needed from the product (see section Common Goals in 2.2.3 Factors).

Included Trust-Affecting Factors:
Mental Model, Expert/Reputable and Common Goals.

5.2.2 *Explicit Information*

What is happening?

In contrary to implicit information (the first event in the pre-use phase), explicit information is actively chosen, first-hand information, gathered from for instance a vehicle dealer or a vehicle owner. This could be the first time the possible customer gets a view of the physical exterior of the vehicle.

What is needed?

Being able to gather the relevant information in order to understand how the AD system will work in a possible future driving situation.

What affects trust?

As in implicit information*,* it is equally important here to understand that the AD system's intentions, conveyed through for instance a vehicle dealer, also affect the potential customer's mental model (see section Mental Model in 2.2.2 Issues). Once again the system purpose conveyed and the real performance of the system must be one and the same, or else the user could lose his or her trust for the AD system in a future interaction (see section Loss of Trust in 2.2.2 Issues). It should be presented as a very good system if the performance allows it (see section Expert/Reputable in 2.2.3 Factors). It could also be favorable to express how the common goals are similar between the user and the AD system (created by the company's designers) in order to create trust (see section Common Goals in 2.2.3 Factors). The last part during this event is the importance of having pre-training, i.e. instructions presented by the vehicle dealer or an actual training program, in order to teach and familiarize the user with the system and how it works, focusing on the system's functions rather than future possible situations (see section Training in 2.2.3 Factors). Focusing on functions is favorable since it allows the user to learn how to search and find information in the AD system rather than to just learn how to act in a certain situation.

Included Trust-Affecting Factors:
Mental Model, Expert/Reputable, Common Goals and Training.

5.3 Learning Phase

5.3.1 *Enter Vehicle*

What is happening?

During this event the user steps into the vehicle for the first time, getting an initial impression of the interior. The user is able to make personal and pre-activation adjustments such as seat, steering wheel and mirror positions, etc.

What is needed?

Be able to enter vehicle and to understand and find adjustment controls in order to change personal settings.

What affects trust?

During this event an anthropomorphic feature that politely welcomes the driver could be advantageous in order to create an appropriate level of trust as well as a greater acceptance for the AD system (see section Anthropomorphism in 2.2.3 Factors). If an anthropomorphic feature is used in the AD system it could be presented for the first time during this event even if the AD system is not engaged, i.e. it can be used as an overall feature connected to the AD system.

The physical aesthetics of the interior also affect how the system is perceived and thereby alter the user's level of trust for the system e.g. a system perceived as competent feels more trustworthy, something that regards all systems (see section Expert/Reputable in 2.2.3 Factors).

Included Trust-Affecting Factors:
Anthropomorphism and Expert/Reputable.

5.3.2 *Activating Vehicle*

What is happening?

The driver activates and checks different systems needed, i.e. not only the AD system. He or she gets the first interior system impression.

What is needed?

To understand how to activate different systems and to interpret the information received from them, understanding system status and what it means (see section 2.1.3 Usability).

What affects trust?

The system should be able to adapt to the driver's existing mental model and preferences, either through pre-existing knowledge about the driver that has been incorporated into the system, or through letting the system scan the driver and by this get the information of what the driver wants and needs (see section Adaptive Automation in 2.2.3 Factors). It should at least be possible to choose individual preferences such as how information should be presented to optimally fit the driver (see section Customization in 2.2.3 Factors).

Presenting the AD system with a personality, e.g. name, gender and voice (which could have already been briefly introduced when entering the vehicle) is preferable to make it more anthropomorphic. These features (in total) should also give the system a look and feel of competence (see sections Anthropomorphism & Expert/Reputable in 2.2.3 Factors).

It is important to get learning feedback through either audio, haptic or visual cues, or all of the above (see section Feedback in 2.2.3 Factors). This learning feedback should focus on system functions, or even have AD training that could be conducted by the system focusing on the functions rather than plausible situations in order to create an appropriate trust for the system (see section Training in 2.2.3 Factors). This could also help create an appropriate level of trust by familiarizing the user with the system.

Included Trust-Affecting Factors:
Expert/Reputable, Customization, Anthropomorphism, Adaptive Automation and Feedback.

5.3.3 *Manual Mode 1*

What is happening?

The driver gets the first impression of vehicle driving functions. This mode is when driving on uncertified roads (i.e. roads not approved for AD systems).

What is needed?

A basic set of manual driving knowledge is needed. Be able to understand and interpret the environmental situation, traffic regulations and system information (see section 2.1.3 Usability). React to the input within a sufficient time depending on the situation. Understand where you can use the AD system as well as know where to find the AD-certified roads.

What affects trust?

The AD system's anthropomorphic features should be presented in a non-intrusive and non-annoying way (see section Anthropomorphism in 2.2.3 Factors). Information could be given of where certified roads can be found. It could also be favorable to be able to change non-critical settings such as different types of system information output for instance the voice of the anthropomorphic feature (see section Customization in 2.2.3 Factors).

The adaptive system steps in and assists when needed i.e. critical situation interventions, braking, steering etc. It is also highly important that feedback is given during these incidents so that the driver understand the system's actions (see sections Adaptive Automation & Feedback in 2.2.3 Factors).

Included Trust-Affecting Factors:
Customization, Anthropomorphism, Adaptive Automation and Feedback.

5.3.4 *Control Transition 1*

What is happening?

Driving system handover, i.e. from manual driving the vehicle to letting the AD system take over the controls. This only occurs on certified roads.

What is needed?

The driver needs to understand how to activate the AD system as well understand if the current environmental situation allows for a control transition. He or she also needs to understand possible feedback from the system regarding activation and by that know who is in control and have an understanding of the transition procedure (see section 2.1.2 Mode Confusion).

What affects trust?

The AD system should be able to adaptively decide if a control transition (CT) is possible or at least give information of what is favorable from a safety perspective (see section Adaptive Automation in 2.2.3 Factors). If a CT is possible, common goals can be used e.g. different driving styles such as eco-mode, smooth mode or fast mode from which the driver can choose, and such goals could be presented through the anthropomorphic feature of the AD system (see sections Common Goals & Anthropomorphism in 2.2.3 Factors). This feature should allow a non-interruptive and patient feedback, but still be portrayed as competent (see section Expert/Reputable in 2.2.3 Factors). The driver could also be able to actively customize different settings, but only those that will never affect safety/critical information because this could lead to driver confusion (see section Customization in 2.2.3 Factors). These settings could be how the anthropomorphic feature should be portrayed so it does not become annoying for the driver. In safety critical situations like control transitions it is important to show system intentions through both how and why information (see section Why & How in 2.2.3 Factors). An example of why information could be that the AD system communicates that a CT is in progress because the vehicle is on an AD-certified road, and how information could be that the system is now doing so by first taking over the steering control and then the throttle and braking functions. How and why information could together create driver anxiety, but presenting both is favorable in more critical situations, such as a control transitions, since it is better from a safety perspective as it allows the driver to fully understand what is happening.

Feedback during a CT should be presented very clearly in different ways (audio, haptic & visual), where the visual information should preferably be presented with icons instead of text in order to minimize the already possible high mental workload due to the CT itself (see section Feedback in 2.2.3 Factors). Vocal information is the least cognitively demanding and the easiest to comprehend, but optimally haptic and visual information could also be used together with vocal information. Uncertainty information should also be presented if the system is "unsure" if a CT is recommended in the current situation e.g. if the sensors are not fully recognizing the environment (see section Uncertainty Information in 2.2.3 Factors). The AD information should always be more prominent than other information e.g. knowing that the system is on or off as well who is in charge, system or driver (see section 2.1.2 Mode Confusion). It is also important to focus on intuitive placements for certain controls e.g. AD driving controls could be placed on or close to the steering wheel since they affect driving, particularly if they are not very prominent (see section 4.6 General Thoughts in 4 User Study).

Included Trust-Affecting Factors Included Trust-Affecting Factors:
Expert/Reputable, Customization, Anthropomorphism, Adaptive Automation, Common Goals, Why & How Information, Uncertainty Information & Feedback.

5.3.5 *Autonomous Mode*

What is happening?

The AD system is engaged and in full control over the vehicle (level 3 NHTSA).

What is needed?

The things a driver needs to understand during AD mode is the functionality of the mode, i.e. understand fully what AD mode does and does not do as well as understand when and where to put attention on what; on the AD system feedback or on the environment or on both (see section 2.1.3 Usability & 2.1.2 Mode Confusion). It is also important to know where to find relevant system feedback as well as be able to interpret it. The driver also needs a moderate alertness in case of a control transition (see section 2.1.1 SA – Situation Awareness).

What affects trust?

During AD mode it is important as a driver to be able to change the common goals, agreed upon earlier, e.g. shifting from one driving mode to another (see subsections Common Goals & Customization under 2.2.3 Factors).

Continuous information how the common goals are met is also favorable. The adaptive system should present information about workload needed for a possible control transition and also decide if manual mode is recommended and what is needed from the driver, depending on environmental factors (see section Adaptive Automation in 2.2.3 Factors).

In AD mode the system could be given a higher mental state through human abilities and an expert expression, for instance conveying information through accented, although some accents could be perceived as more expert-like than others (see section Expert/Reputable & Anthropomorphism in 2.2.3 Factors).

Non-critical information output should be customizable, but critical information output should stay the same. The non-critical information could be environmental feedback as well as suggestions about certain actions that could be taken, for instance overtaking or a lane change (see section Feedback in 2.2.3 Factors). The system should show intention by only presenting information about why it is performing certain actions (i.e. braking for a vehicle ahead) since both why and how information can generate a high mental workload and cause anxiety for the user in the long run (see section Why & How in 2.2.3 Factors). During AD mode, the system should convey any uncertainty since it helps create an appropriate level of trust by showing the system's limitations, which also increases situation awareness (see section Uncertainty Information in 2.2.3 Factors).

The feedback during AD mode should be continuous and present information about upcoming events as early as possible in a way that allows the driver to focus on the system and on the environment at the same time (see section 4.6 General Thoughts in chapter 4 User Study). Feedback could be good to present through vocal information but this could also become annoying, therefore being able to customize or use different information outputs (visual, vocal, audio & haptic) together could be more optimal.

Included Trust-Affecting Factors:
Expert/Reputable, Customization, Anthropomorphism, Adaptive Automation, Common Goals, Uncertainty Information, Why & How Information & Feedback.

5.3.6 *Control Transition 2*

What is happening?

A second driving system handover (in reverse order to CT1). The control of the vehicle shifts from the AD system to the driver.

What is needed?

The driver needs to understand how the control transition procedure will be executed, when and how to deactivate the AD system as well as have a high level of situation awareness (SA) in order to swiftly adjust to the new driving situation by knowing where to focus and understanding the environment. It is also important to allow the driver to fully understand when he or she is in control again (see sections 2.1.3 Usability, 2.1.1 SA – Situation Awareness & 2.1.2 Mode Confusion).

What affects trust?

An adaptive system should advise through status information on how the current environment looks (see section Adaptive Automation in 2.2.3 Factors). The communication is mediated politely through the competent anthropomorphic part of the system (see sections Anthropomorphism & Expert/Reputable in 2.2.3 Factors). Customization settings will never override safety/critical information as this could lead to driver confusion. In safety critical situations like control transitions it is important to show system intentions through both why and how information e.g. showing how the CT process will be carried out (see sections Customization & Why & How Information in 2.2.3 Factors).

During the CT, feedback is very important since it builds trust, and the information should be presented in different ways (audio, haptic & visual) in order to minimize the already high mental workload because of the CT itself. Vocal information is the less cognitively demanding and the easiest to comprehend, but optimally haptic and visual information could also be used (see section Feedback in 2.2.3 Factors).

Included Trust-Affecting Factors:
Expert/Reputable, Customization, Anthropomorphism, Adaptive Automation, Why & How Information & Feedback.

5.3.7 *Manual Mode 2*

What is happening?
Driving manually on uncertified roads. Driving to and arriving at the destination.

What is needed?
After the control transition it is necessary to have manual driving alertness, since the workload and context have changed from AD control to manual drive, which means that a cognitive confusion could arise (see section 2.1.4 Loss of Skill).

What affects trust?
During this event it is important to have an adaptive system that steps in and assists when needed e.g. critical situation interventions such as braking assist and similar help systems (see section Adaptive Automation in 2.2.3 Factors).

Anthropomorphic features are always present in a non-intrusive and non-annoying way (see section Anthropomorphism in 2.2.3 Factors). It is also important that feedback is given during these situations (see section Feedback in 2.2.3 Factors). Being able to change non-critical settings such as different types of information output regarding other systems also instills trust (see section Customization in 2.2.3 Factors). In this event, it could also be beneficial to show how the earlier set common goals are met, e.g. showing how efficient eco-mode has been during AD mode (see section Common Goals in 2.2.3 Factors).

Included Trust-Affecting Factors:
Customization, Anthropomorphism, Adaptive Automation, Common Goals & Feedback.

5.3.8 *Exit Vehicle*

What is happening?
Turning off all systems and leaving the vehicle.

What is needed?
Understand how to get vehicle into stationary status and be able to deactivate the different systems.

What affects trust?
When exiting the vehicle, a polite anthropomorphic system could be beneficial in order to create trust. This could be done through a goodbye phrase (see section Anthropomorphism in 2.2.3 Factors).

Included Trust-Affecting Factors:
Anthropomorphism.

5.4 Performance Phase

5.4.1 *Continuous Usage*

What is happening?

This event explains how a continuous usage of the AD system would look *without* a change or an incident.

What is needed?

Retain the knowledge obtained in the learning phase as well as manual driving skills.

What affects trust?

Later in the human-machine relationship, trust is mostly based on the AD system's dependability, i.e., it is not as important to show intention as is getting feedback about the AD system's performance. The adaptive system has learned the driver's preferences and has adapted to them, e.g. preferred driving style (see section Adaptive Automation in 2.2.3 Factors). Also customization is needed since the driver could find certain information to be annoying after a while and could therefore actively choose if it should be presented or not, or reduce the level of guidance by decreasing the level of information (feedback) and how it is presented (see section Customization & Feedback in 2.2.3 Factors).

Included Trust-Affecting Factors:
Adaptive Automation, Customization & Feedback.

5.4.2 *Change of Context*

What is happening?

This event covers changes in environment, behavior, or system, e.g. new roads, behavioral changes, or smaller system updates.

What is needed?

If a change of context occurs, the user's knowledge about how the system reacts to new environments needs to be updated. Ideally the user should be aware that confidence can be affected by internal or external factors, potentially affecting one's behavior and trust, which the HMI may compensate for by increasing SA via uncertainty information. The driver also needs to understand system updates.

What affects trust?

In a change of context the driver may feel more anxious and therefore it could be good to have a more transparent system, through showing more intention until the level of trust has been stabilized again (see section Feedback in 2.2.3 Factors). An adaptive system should continue to adapt to new changes, e.g. behavioral changes. If the environment is new it could be optimal to show a higher degree of uncertainty information (see section Uncertainty Information in 2.2.3 Factors).

Included Trust-Affecting Factors:
Feedback & Uncertainty Information.

5.4.3 *Incident*

What is happening?

Incident in AD mode or in control transitions.

What is needed?

A high alertness as well as an understanding about what is necessary in a critical situations and existing knowledge of what the system is capable of and of how to react.

What affects trust?

The system decreases the number of incidents through assisting the driver in failure response and detection. Users have a tendency to mitigate less blame on anthropomorphic systems in case of incidents (see section Anthropomorphism in 2.2.3 Factors). If an incident occurs because of human error, it could be beneficial to correct the level of trust through training (see section Training in 2.2.3 Factors).

Error information is also needed after an incident, but it should not focus on obvious errors; rather it should continue to show system performance as well as explain how the errors will affect the overall system performance. It is especially crucial to show error information in the learning phase (see sections Error Information & Feedback in 2.2.3 Factors). The information presented, for instance in a braking incident in AD mode, should explain why the system is braking (see section Why & How Information in 2.2.3 Factors).

Included Trust-Affecting Factors:
Anthropomorphism, Feedback, Error Information, Why & How Information & Training.

6 Concept Development

The concept development section describes several iterations made between ideation, compilation and selection.

6.1 Ideation

The two workshops with engineering students (see figure 26) and experts from Volvo Car Corporation generated around 300 ideas. These ideas had to be compiled and developed, by first collecting them and putting similar ideas together in order to make the volume more manageable.

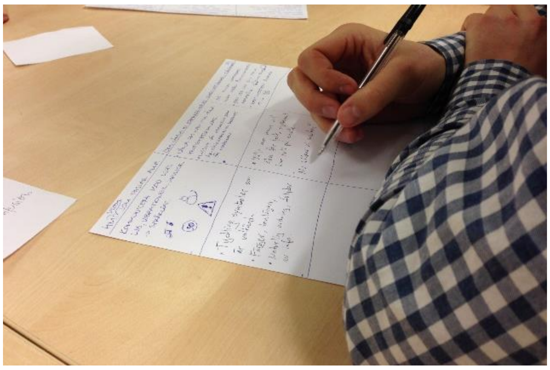


Figure 26 - Workshop with engineering students at Chalmers University of Technology.

The concept development process was highly iterative, going from compiling the ideation results, to developing them further and back to compiling them again (see figure 27). The iterative process increased the potential of the final result since it is based on a continuous shift between synthesis and analysis (Johannesson, et al., 2004), i.e., creating better and better concepts through every iteration loop. Due to the divergent complexity of the task at hand (because of the many different trust components that needed to be connected to the HMI concepts), the result from the ideation had to first be extracted, compiled and clustered through a two-step affinity diagram.

6.2 Compilation

During the compilation the around 300 different ideas were gathered into different themes with the help of two affinity diagrams, here called initial and final analysis respectively. These diagrams enabled a structured compilation process as well as facilitated a common ground for communication based on themes or classifications which are easier to grasp (Chalmers University, 2005).

In the initial analysis, the approximately 300 different ideas became 22 themes in total, ranging from Anthropomorphism to Quality (see table 2).

Figure 27 - Compilation of Ideas.

Table 2 - Initial Analysis

Anthropomorphism	*Feedback Placement*
Expert	*Error Information*
Adaption	*Openness*
Customization	*Mental Model*
Two-way Communication	*Gamification*
Transparency	*Simplicity*
Safety Functions	*Aesthetics*
Training	*Familiarity*
Senses	*Soft Transition*
System Understanding	*Friendly*
In-the-loop	*Quality*

During the final analysis the 22 themes from the initial analysis were combined once more (combining old themes into new ones), decreasing the amount to 10 themes. (See table 3).

Table 3 - Final Analysis

Insight	*Communication*
Adaption/Customization	*Common Goals*
Where to Present	*How to Present*
Learning	*Transition*
Competent	*Anthropomorphism*

The themes were then illustrated further creating more elaborate ideas, which were then put into a total of ten different matrices. In these matrices the ideas were organized to be able to more easily visualize the ideas (see figure 28). After they were organized, some ideas were drawn from the matrices and put into one of three different main HMI concepts (see section 6.3 Main HMI Concepts), based on how well the ideas were thought to work together, hence creating three different main concepts.

Figure 28 - Matrices.

6.3 Main HMI Concepts

In this section the three main HMI concepts will be presented. These main concepts are based on the earlier described ideas. The three main concepts focus on placing the different ideas in the driver environment so as to create an appropriate level of trust for the AD system. Presented below are the prominent features from the different main concepts. More detailed information about each concept be found in the following sections (see sections 6.3.1-6.3.3)

Concept one (see figure 29) is characterized by a highly anthropomorphic function creating a feeling of a more human-to-human interaction. This system allows the driver to get a lot of information through the "shadow man" anthropomorphic feature that appears when certain information needs to be conveyed or the driver needs assistance.

Concept two (see figure 30) is mostly characterized by its simplicity and the reduction of complexity. The main features are the visual and simplistic design of the interfaces and menus as well as the absence of the gearshift, which has been replaced with a touch pad.

The most prominent feature in the third concept (see figure 31) is that it lacks a classic DIM (Driver Information Module). Instead it uses the windscreen and the steering wheel as plattforms to mediate necessary driving information.

6.3.1 *Concept One*

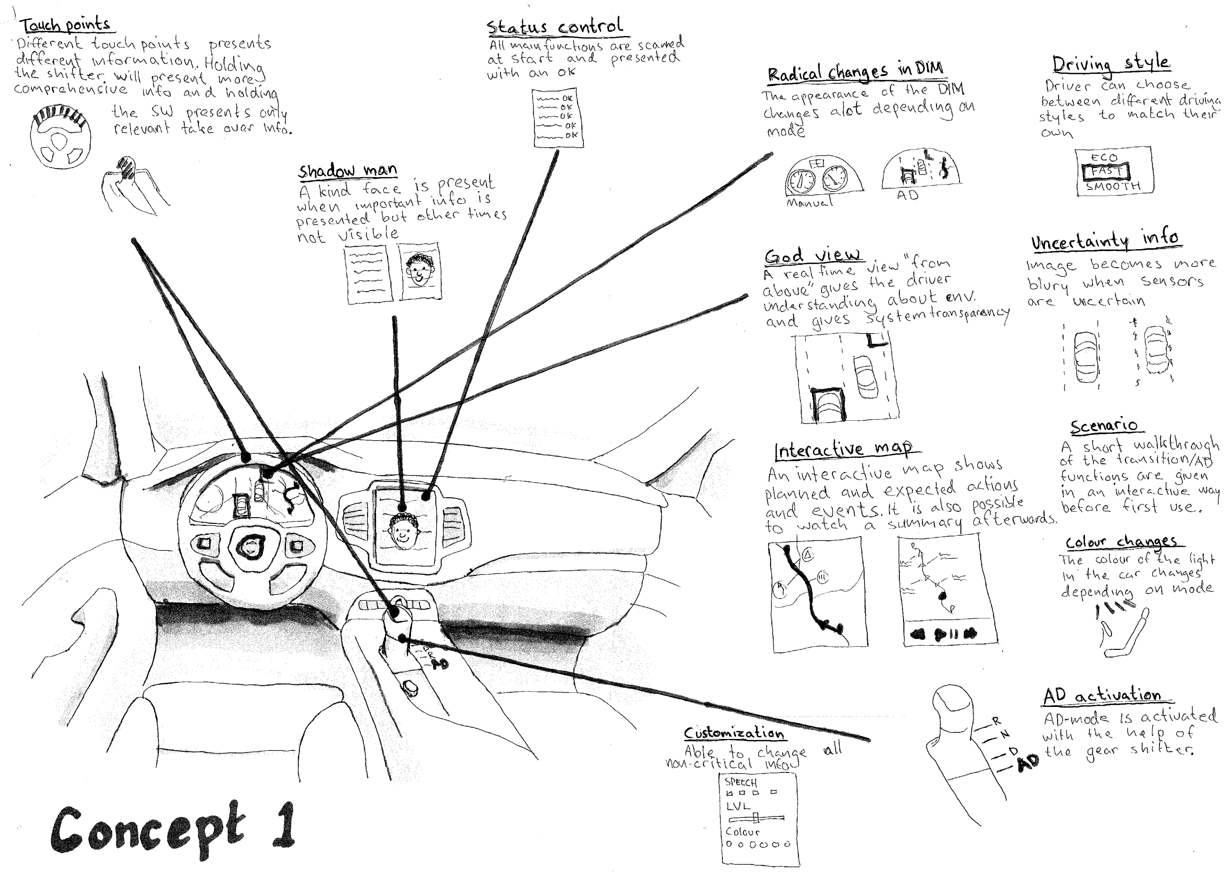


Figure 29 - Concept One.

Touch Points

The system uses different touch points, i.e. different information will be presented depending on where you put your hands during AD-mode.

With the hands on the steering wheel, the driver will only get information that is relevant regarding takeover situations, i.e. important driving information. It could be feedback such as how the autonomous driving system is coping with the environment, e.g. getting object recognition information allowing the driver to understand how the system is processing the environmental information. The god-view function as well as the uncertainty information could be closely related to the feedback given when holding the steering wheel during AD mode.

If the driver put his or her hands on the gearshift, more comprehensive information will be presented, i.e. all other information such as information about the route (interactive map), infotainment settings, climate control settings as well as time to control transition etc.

Included Trust- and HMI-Affecting Factors: Feedback & Situation Awareness.

Shadow Man

The shadow man function is a highly anthropomorphic feature that is very human-like and uses speech as well as has a human appearance. This shadow man is polite and only appears in order to give feedback to the driver. The feature cannot be intrusive and should only appear for certain information exchanges. It could present itself in the beginning by welcoming the driver, and could be more prominent during the control transitions and during AD mode. The shadow man could be more important during the learning phase. The driver should be able to decrease this feature if desired.

Included Trust- and HMI-Affecting Factors: Anthropomorphism.

Status Control

When starting the vehicle and activating the system, the system shows a system check. This system check or status control allows the driver to see that the system is processing and giving feedback about the current system status.

Included Trust- and HMI-Affecting Factors: Expert & Reputable.

Radical changes in DIM (Driver Information Module)

Radical changes in the DIM means that when a control transition is happening, the DIM changes to create a totally different visual experience as well as shows other information during AD mode compared to manual mode.

Included Trust- and HMI-Affecting Factors: Mode Confusion.

Driving Style

The driver is able to choose between different driving styles that best fit to his or her own preferences. Examples could be smooth, fast or eco driving mode.

Included Trust- and HMI-Affecting Factors: Common Goals.

God View

The god view feature gives the driver a real-time top view showing the environment and objects surrounding the vehicle. This allows the driver to understand what is happening around him or her as well as creates a system understanding, i.e. an understanding of what the system is recognizing.

Included Trust- and HMI-Affecting Factors: Feedback & Situation Awareness.

Uncertainty Information

Uncertainty information means that when the environmental sensors become “unsure”, e.g. in a highly complex context, the visual image presented of the environment becomes more and more blurry, which gives the driver feedback about the system’s current status and that it could be good to get ready for a possible control transition from the AD mode to manual mode.

Included Trust- and HMI-Affecting Factors: Uncertainty Information & Feedback.

Interactive Map

An interactive map feature could be favorable for the driver in order to see where possible control transitions will be executed, for instance, going from the highway to a more complex environmental context. This could also be good for system understanding since the driver will get an enhanced understanding about upcoming actions. This function is connected with the system’s GPS (Global Positioning System) and constantly changes depending on new system information. Afterwards, the driver could get a summary on what happened during AD mode.

Included Trust- and HMI-Affecting Factors: Situation Awareness, Common Goals & Feedback.

Scenario

An interactive walkthrough is given by the AD system before first time usage. This walkthrough allows the driver to learn important functions as well as learn how the AD system will give feedback to the driver. The driver is also able to try the different functions within the system before even starting the vehicle.

Included Trust- and HMI-Affecting Factors: Training.

Color Changes

There are two different mode colors in the vehicle: one for AD mode and one for manual mode. The colors change from one to the other during a control transition.

Included Trust- and HMI-Affecting Factors: Mode Confusion.

Function Customization

Customization means that the driver has the possibility to change all non-critical information, e.g. the type of information output as well as the level of information output. An example could be changing the anthropomorphic feature from on to off or choosing how much this function should be active during non-critical situations such as manual mode, or turning off hello messages when stepping into the vehicle.

Included Trust- and HMI-Affecting Factors: Customization.

AD Activation

The AD mode is activated through the gearshift, similar to how driver would choose D for drive or R for reverse.

Included Trust- and HMI-Affecting Factors: Usability and Feedback

6.3.2 *Concept Two*

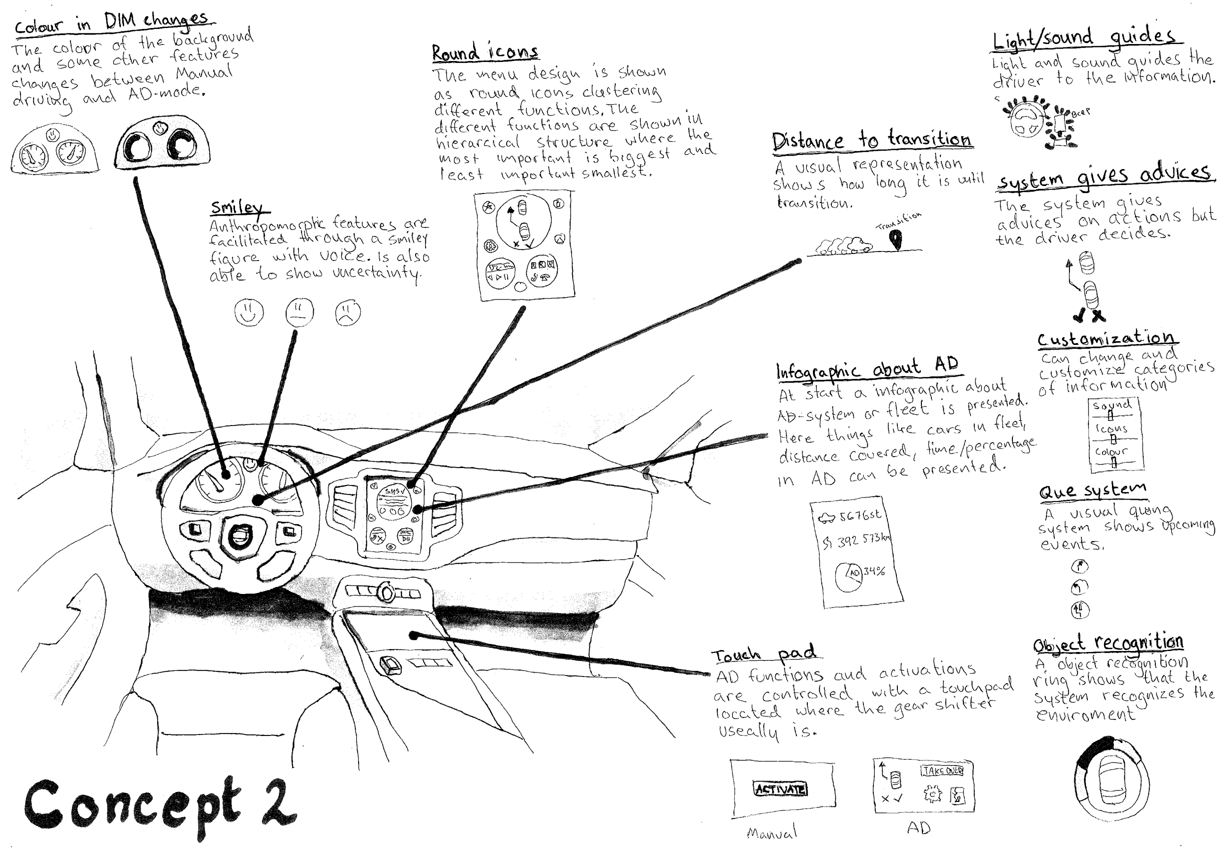


Figure 30 - Concept Two.

Color Changes in the DIM (Driver Information Module)

The color of the background and of some other features, such as the speedometer, changes in the DIM between manual mode and AD mode. This is to enhance awareness of which mode is active to minimize mode confusion.

Included Trust- and HMI-Affecting Factors: Mode confusion.

Smiley

Anthropomorphic features are facilitated through a smiley face with voice. It could be more prominent during the control transitions and during AD mode, vocally informing the driver about important information. The feature can also show uncertainty information by showing different smiley symbols with different facial expressions depending on system status.

Included Trust- and HMI-Affecting Factors: Anthropomorphism.

Round Icons

The menu design is shown as round symbols clustered according to different functions. The functions are shown in a hierarchical structure where the most important information is represented by the biggest symbols and the least important information by the smallest. Examples of different functions could be AD information or various infotainment controls.

Included Trust- and HMI-Affecting Factors: Usability, Expert/Reputable

Distance to Transition

In order to prepare the driver for a control transition a visual representation shows approximately the distance to a possible handover. A car figure moves towards a transition mark to give the driver time to prepare for the take over.

Included Trust- and HMI-Affecting Factors: Feedback & Workload.

Light/Sound Guides to Information

Light and sound guides the driver to know where to find the relevant information for the specific situatuion. For example if the information is available in the DIM or in CSD (Center Stack Device), the respective cluster lights up and makes a sound.

Included Trust- and HMI-Affecting Factors: Feedback & Usability.

System Gives Advice

The system gives advice on actions but it is up to the driver to make the decision on whether to act upon it or not. This is done to enhance the feeling of being a team which can increase the level of trust.

Included Trust- and HMI-Affecting Factors: Common goals.

Infographic about the AD system

A clear and easily understod infographic is presented at the start, containing information about the AD system. The infographic could present information such as distance traveled, time in AD mode or percentage of time in AD mode without any incidents.

Included Trust- and HMI-Affecting Factors: Expert/Reputable & Feedback.

Function Customization

The concept is customizable to a certain extent, meaning that the driver is able to customize certain categories of non-critical information, but is not able to go in and change the meaning of the information. One example could be that it is possible to adapt the icons with some predefined options but you are not able to change the implication of the icons functions.

Included Trust- and HMI-Affecting Factors: Customization.

Cue System

A visual cueing system shows pending actions so the driver can get an insight in what the system “thinks”, creating system transparency and also an understanding about why the system acts as it does.

Included Trust- and HMI-Affecting Factors: Feedback and Why & How information.

Object Recognition

An object recognition ring shows that the sytem recognizes the surrounding objects. The ring indicates both distance and direction by coloring different sections in orange or yellow.

Included Trust- and HMI-Affecting Factors: Feedback & Situation Awareness.

Touch Pad

AD activation and functions during AD mode are controled with help of the touch pad that is located where the gearshift usually is located. When in manual mode, the touch pad only shows the AD activation button and when entering AD mode the display changes to show functions connected to AD mode.

Included Trust- and HMI-Affecting Factors: Feedback & Mode Confusion.

6.3.3 *Concept Three*

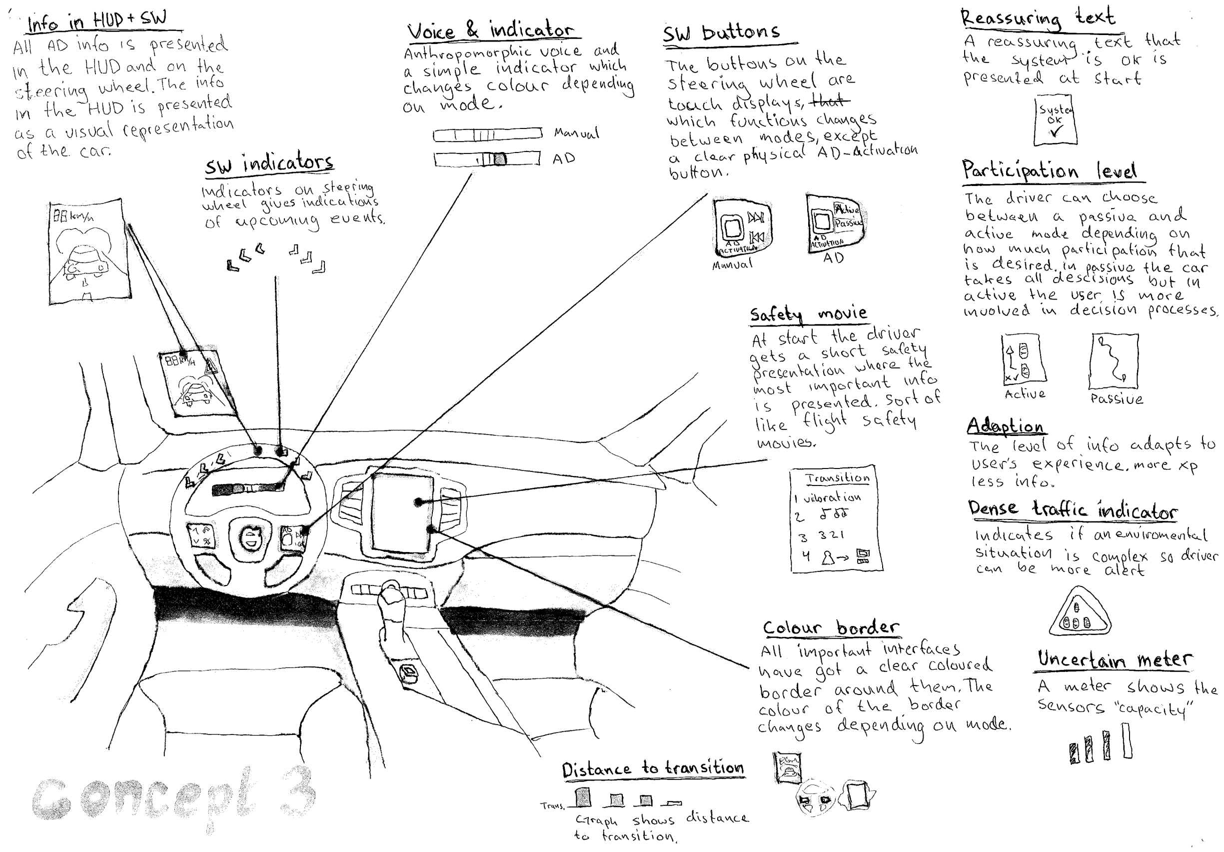


Figure 31 - Concept Three.

Information in HUD + SW

This concept uses the windscreen and/or steering wheel (SW) to convey the proper driving information. The HUD (Head-Up Display) uses an illustration of the vehicle and all the necessary information connected to it instead of using any classically styled speedometer or odometers. The dense traffic indicator, uncertainty meter, as well as the distance to transition are presented here. By presenting everything here, this allows the driver to understand where to find and easily access the necessary information. The steering wheel could also be used as a platform in order to convey this information.

Included Trust- and HMI-Affecting Factors: Usability.

Steering Wheel Indicators (SW)

These are indicators giving information about upcoming take overs.

Included Trust- and HMI-Affecting Factors: Feedback.

Voice Indicator

This concepts uses a voice in order to convey information and to allow a more human-like personality. This voice could welcome the driver as well as mediate information during certain events. The indicator is used to only show if the system is in AD mode or in manual mode.

Included Trust- and HMI-Affecting Factors: Anthropomorphism & Mode Confusion.

Steering Wheel Buttons

These are pressure-sensitive displays on the steering wheel which are used as buttons. These button functions change depending on if the system is in AD mode or in manual mode, except for a fixed AD activation button that always stays the same. During manual mode these button functions could be used for volume or channel control. In AD mode they can be used for AD functions such as overtaking, changing speed or swiping between different information outputs in the HUD.

Included Trust- and HMI-Affecting Factors: Usability & Mode Confusion.

Safety Movie

This feature is a safety movie playing before the driver's first usage in order to illustrate how the AD mode works and its connected functions.

Included Trust- and HMI-Affecting Factors: Training.

Reassuring Text

When activating the vehicle a text presents itself conveying the system status. This is only presented by a "System OK" or an error message explaining what is wrong.

Included Trust- and HMI-Affecting Factors: Expert & Reputable.

Participation Level

It is possible for the driver to shift between two different AD modes: one more driver active and one more driver passive. The active mode allows the driver to choose if the vehicle should do a takeover and decide a preferable speed. When the passive mode is chosen, the AD system decides the actions for the driver. When the driver for instance chooses to overtake within the passive mode, the AD mode changes from passive to active. If the driver has active mode on but is not doing anything, the system will after a period of time go back to passive mode. The difference between the modes is mirrored in how information is presented; more information during active mode and a more general route perspective during passive mode.

Included Trust- and HMI-Affecting Factors: Common Goals.

System Adaptation

The information level changes depending on the driver's experience, hence more information in the beginning of the usage (learning phase) compared to later.

Included Trust- and HMI-Affecting Factors: Adaptive Automation.

Dense Traffic Indicator

Indicates if an environmental situation is complex so that the driver can be more alert.

Included Trust- and HMI-Affecting Factors: Feedback & Situation Awareness.

Uncertainty Meter

A meter shows how "sure" the different sensors are about the environmental context, as well as if they are fully functional.

Included Trust- and HMI-Affecting Factors: Uncertainty Information & Feedback.

Color Border

All important interfaces have a lit border around them. This border changes color depending on which mode is active, manual mode or AD mode.

Included Trust- and HMI-Affecting Factors: Mode Confusion.

Distance to Transition

A graph shows the distance to an upcoming control transition.

Included Trust- and HMI-Affecting Factors: Feedback & Workload.

7 Validation Test

This chapter describes the results from the final validation test. This test was a way to evaluate the three different object recognition concepts (OR-concepts) from the three main HMI concepts presented in sections 6.3.1-6.3.3. The methods used are described in section 3.6 Validate.

7.1 OR-Concepts

The three designed OR-concepts have different levels of system transparency ranging from just showing that an object is near the vehicle to showing more exactly what kind of object there is and where it is in relation to the car. The OR-concepts are called Sensor Indication Concept, Ring Concept and God-View Concept.

Sensor Indication Concept is the concept that gives the least system transparency of the three. It consists of a round symbol showing when an object is within close range of the car. The symbol can be seen in figure 32.

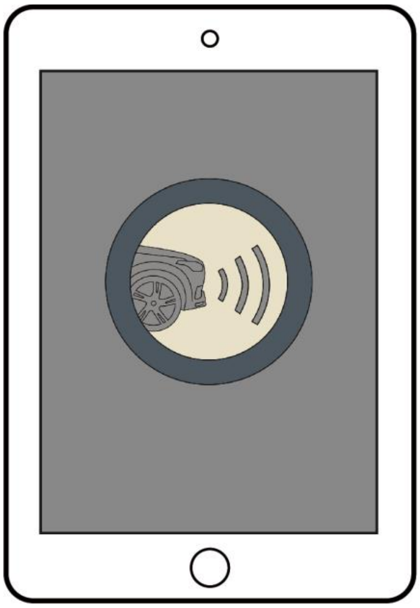


Figure 32 Sensor Indication Concept interface, shown on a tablet as used during the validation test.

The Ring Concept shows system recognition through a representation of the car with a light grey ring around it. This grey ring can show direction and distance to objects by having different parts of the ring change color to either yellow or orange depending on distance to object (see figure 33).

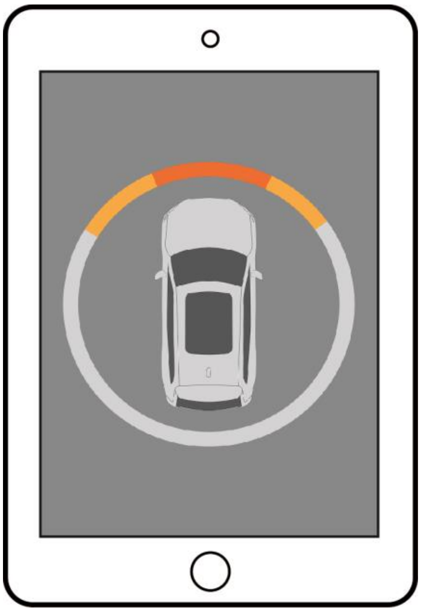


Figure 33 - Ring Concept interface, shown on a tablet as used during the validation test.

God-View Concept is the one that conveys most system transparency by also showing object type. This concept is visualized through a representation of the car. It also shows objects around the car (see figure 34).

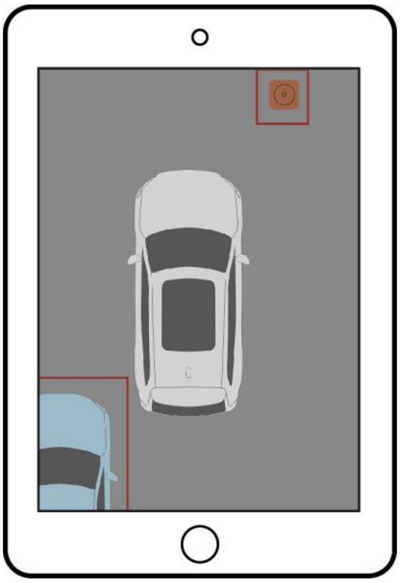


Figure 34 - God View Concept interface, shown on a tablet as used during the validation test.

7.2 Result

Participant trust for each concept was measured using a questionnaire consisting of seven questions answered on a 7-point scale (where 1 is "I totally disagree" and 7 is "I totally agree") developed by Helldin, et al. (2013). The results can be seen in figure 35. The results show that the highest average score was for the Ring Concept (4.59) and God-View Concept (4.59), followed by the Sensor Indication Concept (3.46) and finally the control lap with no concept (2.50).

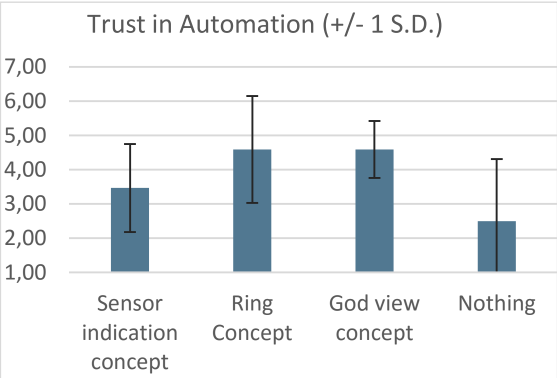


Figure 35 - Mean results from the "trust in automation" questionnaire.

A non-parametric Friedman test (α=5%) found a statistically significant difference in levels of trust between groups, $\chi^2(3)=16.385$, $p=0.001$. Post hoc analysis with Wilcoxon signed-rank tests were conducted with a Bonferroni-Holm correction applied to control for Type I error. At α=5%, no significant differences were found, however, at α=10%, significant differences were found for god view – sensor indicator (Z=-2.524, p=0.012, threshold p<0.017), and god view – nothing (Z=-2.521, p=0.012, threshold p<0.02). One may also argue against a correction due to a low sample size as a study limitation, in which case statistically significant differences are also found for ring – sensor indicator (Z=-2.103, p=0.035) and ring – nothing (Z=-2.100, p=0.036), as could easily be expected from the trust ratings presented above. In either case, the authors recommend a follow-up study with a larger sample size to strengthen the statistical results.

In the interviews, several of the participants stated that the Ring Concept was the most favorable since it is very intuitive. Some said that they did not need to know exactly what the system identifies, as long as they know that it can see the obstacles. However, one participant did not fully understand how the Ring Concept worked.

The God-View Concept was believed to be especially good for orientation purposes since one could identify certain objects on the screen and then compare them with the outside environment to get an overview of the situation. One participant also commented that it could be favorable to know what kind of object that is identified, since some objects are static and some are dynamic.

The general opinion regarding the Sensor Indication Concept was that it gave too little feedback, especially since it does not indicate object placement in reference to the car. Some participants even said they became more stressed by it.

In the lap with no concept, participants stated they felt uncomfortable not having any feedback and that it felt like anything could happen. One participant also said "even if you trust the system, you still want to have some kind of information".

If an error would occur and the system would make a mistake in the visualization, many believed that the driver would be more forgiving of the Sensor Indication Concept and the Ring Concept than they would be of the God-View Concept.

7.3 Analysis

The results from the questionnaire show that the God-View Concept and the Ring Concept generate a higher level of trust compared to not being presented with any feedback at all (both concepts got a mean score of 4.59 compared to 2.5 without feedback). However, in the interviews, several participants identified the Ring Concept as the most favorable since it was intuitive and the most forgiving concept of them all.

Interview results also show that drivers presented with no feedback felt uncomfortable and that the Sensor Indication Concept did not give enough feedback, probably because it lacked information about object placement in reference to the car.

These results suggest that presenting feedback through object recognition, creating a more transparent system, could increase the level of trust for the system. Although, only indicating a nearby object without any feedback about object placement in reference to the car is insufficient.

8 Example Concept

This section presents the final example HMI concept as an illustration of what types of results the framework can help to create. The example concept is based on the three main HMI concepts discussed earlier in section 6.3 Main Concepts. The final example concept should only be seen as an illustrative example of a human-machine interaction concept.

The final example HMI concept is only used to show how a human-machine interaction system can be developed with the help of the framework. The presented result is more of a communicative aid in order to convey the full purpose of the framework and therefore the example concept's ideas need to be further tested, and by this also validate the placement of the different identified trust factors in the different events.

As mentioned in chapter 3 Methodology & Implementation, section 3.7 Present, the pre-use phase is not illustrated here since the aim was to present an HMI concept addressing the direct physical interaction between the AD system/autonomous vehicle and the driver. Even if the pre-use phase is not illustrated here, it is of equal importance. The sub-concepts (on which the example concept is based) are only connected to the events within the learning phase and the performance phase, from 'entering vehicle' to 'incident' (see figure 36).

Figure 36 - Events affected by the example concept.

8.1 Learning Phase

When entering the vehicle a voice welcomes the driver. At the same time facial features appear in the Center Stack Device (CSD) (see figure 37), i.e. the biggest display in the middle of the dashboard. The anthropomorphic features in the CSD recognize the driver by showing that he or she has been noticed. Anthropomorphism (both the voice welcoming the driver and the facial features) creates a greater level of user trust via creating a polite and human-like system.

Figure 37 - Anthropomorphic features shown in the CSD

When activating the vehicle for the first time an instructional video is shown after a system check is performed (see figure 38). The system check is done every time the vehicle is activated, showing the driver that all systems (and especially the AD system) are functional. This system check is used to convey an expert system expression, allowing the user to achieve a higher level of trust.

Figure 38 - System Check.

The instructional video (see figure 39) is presented in the CSD and focuses on the control transition procedure and the AD mode. This video illustrates how the functions work and how they should be operated. It also shows how important feedback will be conveyed. This is shown the first time of usage but it is possible to turn this off after the first usage (or turn it on again later).

This instructional video is implemented in order to generate a system understanding allowing the driver to use the AD system correctly without misuse. This can be seen as a short pre-training by giving learning feedback, only focusing on the most crucial functions.

Figure 39 - Instructional Video. 1 Find AD road, 2 Activate with gearshift, 3 Wait for confirmation, 4 Take the hands off the steering wheel.

After the driver has gone through the instructional video, he or she can manually drive away. During this, the GPS (Global Positioning System) that is connected to the AD mode uses an interactive map showing AD-certified roads, i.e. where it is possible to use AD mode and its functions (see figure 40). The map shows the distance as well as the location of where a control transition is possible, i.e. when entering an AD certified road, which will be displayed as an icon on the map. This map also shows were the certified road ends. Since this is a NHTSA level 3 system, the vehicle cannot be driven autonomously everywhere, and therefore the vehicle needs to show where it is possible to use AD mode.

Figure 40 - Map that shows distance and location of a possible control transition.

During control transition one, activating AD mode is done by dragging the gearshift from manual driving mode (D) (see figure 41) to autonomous driving mode (AD) (see figure 42). AD mode is connected to the gearshift as all other driving modes are, e.g. park (P), neutral (N), sequential gear change ($\frac{+}{-}$), reverse (R) and drive (D).

The AD mode activation is connected to the gearshift since this creates a familiarity, i.e. what the driver knows from earlier interactions with automatic gearboxes, thus making it potentially easier to find the function. This is connected to the usability factor that states it should be intuitive and easy to understand how to use a specific function.

Figure 41 - The gearbox is in manual drive mode (D).

Figure 42 - Above the AD mode is active and the vehicle is operating the driving task itself on an AD-certified road (level 3 NHTSA).

The ambient light changes in the cabin of the vehicle when changing from manually driving the car to AD mode (see figures 43 and 44). This is one of several functions that counteract potential mode confusion by letting the driver fully understand which mode is active.

Figure 43 - Interior in manual mode without blue ambient light.

Figure 44 - Interior in AD mode with blue ambient light.

The steering wheel as well as the Driver Information Module (DIM) change during a control transition as well. The steering wheel's functions during manual mode range from controlling the infotainment unit to activating and changing speed for the cruise control (see figure 45). When AD mode is activated, these functions change to focus on settings connected to AD mode, such as driving styles as well as controlling the interactive map (see figure 46). The driving styles range from active to passive, where passive is split into three types: (1) eco-mode which adapts its driving patterns to target low fuel consumption; (2) smooth mode which allows a soft driving style without too many overtaking events and not too many radical changes in speed (i.e. soft acceleration and deceleration). This will probably make the drive from point A to point B take a longer time but it will be more comfortable. The last passive driving typs is (3) fast mode, in which the AD system will minimize the time to the final destination, but the fuel consumption and the ride will be affected. The active driving style (mentioned above) allows the driver to be more involved in the driving, meaning that he or she gets more information about overtaking scenarios, etc., and has to make more active choices during the route. The active mode uses more momentary input from the driver regarding what is to be done. These driving styles should be presented as choices in order to develop a positive connection between the driver and the system by creating common goals.

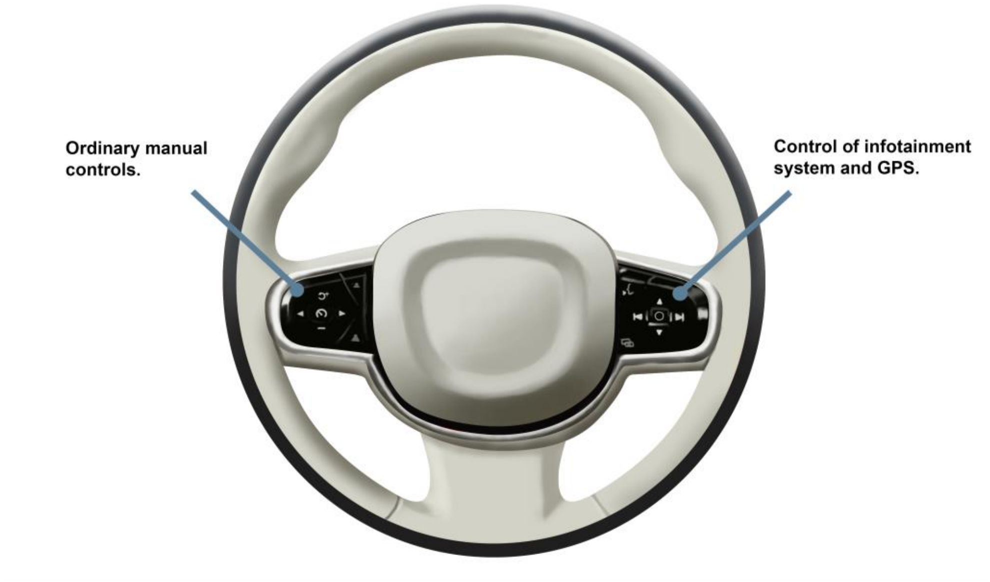


Figure 45 - Steering wheel in Manual mode.

Figure 46 - Steering wheel in AD mode.

The DIM changes radically at the control transitions, showing the traditional gauges in manual mode (see figure 47) and switching completely when entering AD mode, instead showing the interactive map and the object recognition system (see figure 48). The speedometer changes from a traditional gauge to a small digital meter showing the speed. Since the vehicle is autonomously driven, the need for a prominent speedometer is not as crucial. It is important to understand that other types of feedback are more relevant in AD mode than in manual mode and vice versa. The greatest purpose of changing everything during a control transition is however to enable the driver to fully understand which mode is active in order to prevent mode confusion.

Figure 47 - DIM in manual mode showing gauges.

Figure 48 - DIM in AD mode showing interactive map and object recognition system.

Another way of conveying a transition of control, or rather the whole procedure from manual to AD mode, is done by a voice (the anthropomorphic feature) proclaiming relevant information such as distance to transition as well as when the AD system is fully engaged. This information helps the driver to understand how and why these control transition actions are performed.

After the control transition and AD mode is fully active, feedback consists of two major parts presented on the DIM, namely the interactive map and the object recognition system called the Ring Concept (see figure 52).

The interactive map (see figure 49) is a system connected to the GPS system but incorporates relevant functions from the AD system. These functions focus on mediating upcoming actions and environmental events via icons. It can give information about overtaking a vehicle and at the same time present distance to this action (see figure 50). It also presents warning icons (see figure 51) for parts of the highway that have traffic jams or other possible hazardous areas, letting the driver know that his or her attention is now needed. The baseline of the interactive map is early information, making the driver aware of what will happen and leading to system transparency since the driver gets a certain level of understanding of how or rather what the AD system is doing and why it is behaving the way it does.

Figure 49 - Interactive map

Figure 50 - Menu showing distance to upcoming actions and environmental events.

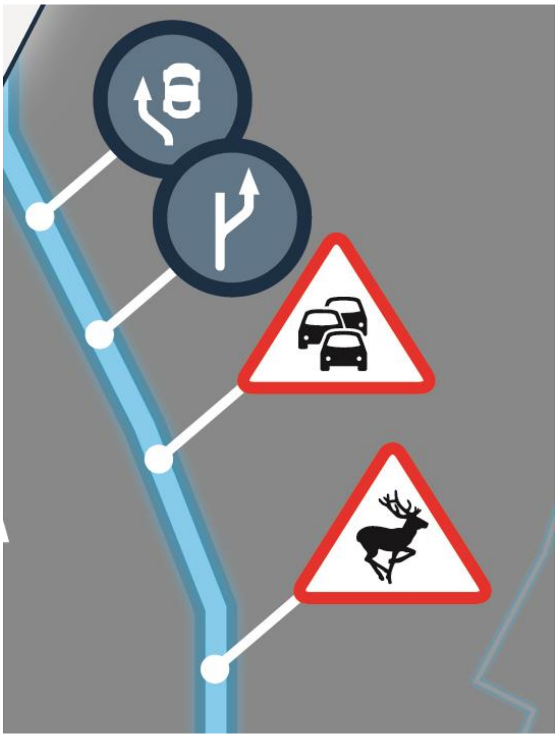


Figure 51 - Icons showing upcoming actions and environmental events.

The interactive map is adaptive in that it constantly updates if new information is available, such as new actions being planned, e.g. takeovers or route changes. Since the HMI system is adaptive, it learns from earlier situations. For instance, if a certain road segment is not optimal for AD mode, the system will adapt to this by giving the driver pre-informational warnings allowing him or her to increase the level of attention during this specific area.

The object recognition system, here called the Ring Concept (see figure 52), indicates distance and direction to obstacles in the exterior environment, such as vehicles and other objects. This is shown by coloring a part of the ring with either yellow or orange depending on how close your vehicle is getting to the obstacle in the environment. The object recognition has a 360° view, so the ring shows objects all around the vehicle. It provides system transparency of what the vehicle "sees" and by that creates an understanding for the driver allowing him or her to achieve a higher level of trust.

Figure 52 - Ring concept

The ring concept also conveys uncertainty information when the AD system sensors are not fully recognizing the environment. It conveys system uncertainty by getting blurry (see figure 53), thus raising the level of awareness of the driver.

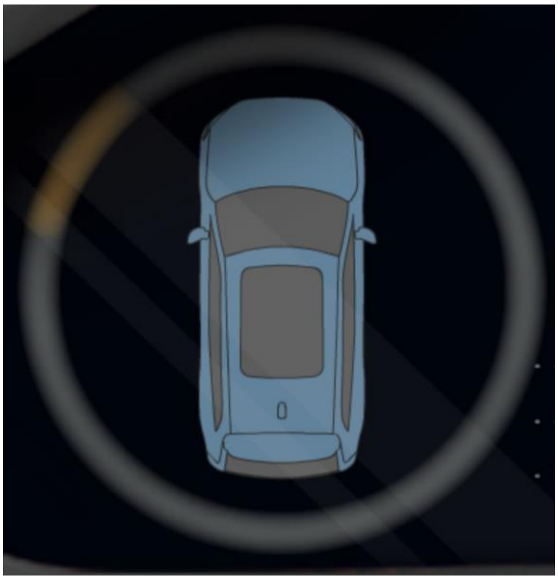


Figure 53 - Ring concept conveying system uncertainty via a more blurry ring.

During AD mode there are also two different touch points: one on the steering wheel and one on the gearshift. These touch points allow the driver to put their hand on either of them. When putting a hand on the gearshift, more in depth information about the interactive map will be shown on the Center Stack Device (CSD) providing a better overview and more comprehensive information of what will happen (see figure 54). The gearshift also has a scroll button to use in order to zoom in and out of the bigger interactive map that appears on the CSD. The scroll button has an 'execute' function as well as 'back' function. These functions can be used for controlling information from the system about for instance upcoming events that can either be denied or executed.

Figure 54 – Touch point on the gearshift showing the interactive map on the CSD.

The touch point on the steering wheel will move the object recognition system Ring Concept to the Head-Up Display (HUD) and a speedometer will appear instead in the DIM (see figure 55). This will happen since the driver wants to quickly get an overview of the situation when taking control, both in case of a forced control transition as well as a planned one. This information is needed in order to raise situation awareness, allowing the driver to get an overview of the environmental context and at the same time see the most important information such as speed and nearby objects. The touch point enables the system to mediate information in a way that is favorable for the driver since it allows him or her to keep the eyes on the road and at the same time get system information that helps raise the level of awareness even more.

Figure 55 – Touch point on the steering wheel moving the ring concept to the HUD and a speedometer will appear instead in the DIM.

During control transition two, the deactivation of the AD mode is done by dragging the gearshift back to drive mode (D) (see figures 56 and 57). It is also possible to take the steering wheel so that the system hands over control to the driver without he or she being forced to put the gearshift into manual drive (D) mode.

The steering wheel touch point is highly important in control transition two since the object recognition system with the Ring Concept moves from the DIM to the HUD, and the speedometer appears in the Ring Concept's previous position on the DIM, will raise situation awareness and will help the driver to prepare for the upcoming manual driving mode.

As in the control transition one, the whole DIM, the buttons on the steering wheel, as well as the ambient light within the driver milieu all change to counteract mode confusion. The control transition is executed in the reverse order of control transition one, affecting the same factors but with a greater focus on situation awareness.

Figure 56 – Interior in AD mode before control transition two.

Figure 57 – Interior in manual mode after control transition two.

After AD mode has been deactivated and the driver is in control again, a summary shows to what extent the common goals that were set for AD mode have been met i.e. if the eco-mode was used it is possible to see how low the fuel consumption was, see the distance and time driven in AD mode, how much carbon dioxide the vehicle has emitted during the AD mode (see figure 58), etc. It is also possible to see more in-depth information about specific events that took place during AD mode. The in-depth information will also be saved for the driver to view later as well as be able to compare this to other routes driven in AD mode, thereby getting a total summary of all trips in AD mode. This function exists to further emphasize the feeling of teamwork.

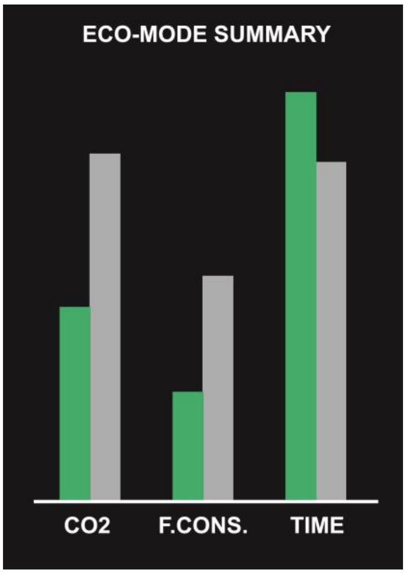


Figure 58 - Summary of the common goals set in AD mode.

When the vehicle has come to a stop, the driver is about to exit the vehicle, and the all the systems have been deactivated, the anthropomorphic facial features and voice reappear (see figure 59). It wishes the driver a good day and hopes the driver has enjoyed the trip. A polite system is perceived as more trustworthy by the driver.

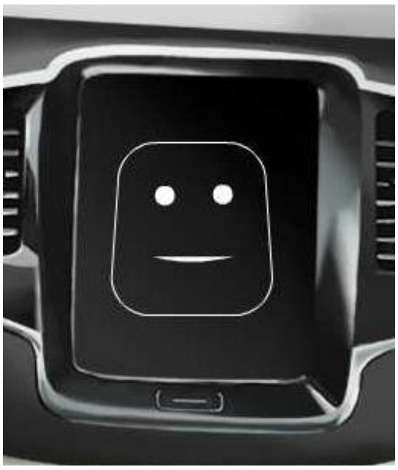


Figure 59 - Anthropomorphic facial features.

8.2 Performance Phase

During continuous usage with the vehicle and its AD system, it is possible to change different settings such as turn the anthropomorphic voice on or off or change the accent.

The HMI system also adapts to earlier preferences, e.g. if the driver often uses eco-mode, this setting will be the default choice for AD mode the driver will have to change it if another driving style is preferred.

Being able to customize (see figure 60) and using an adaptive HMI system minimizes the annoying feeling that can arise when certain information is not needed anymore, i.e. when the driver has learned how certain functions work and the level of trust has now been stabilized relative to the AD system and its performance.

Figure 60 - Showing settings that could be customizable.

If a new context appears such as a new road, the system will automatically increase the level of information for a while until the driver is used to the new environment. One example of information feedback that could increase is information connected to the control transitions. Such information could again become more comprehensive and conveyed in a more anthropomorphic way. This could be done through reinstating the anthropomorphic facial features. This is done as the new environment may cause the driver to feel more anxious, thereby needing information that will once again stabilize the level of trust relative to the AD system and its performance.

Figure 61 - An adaptive system recognizes when the driver needs more information again.

If an incident occurs, such as a very hard deceleration, the system will present error information showing why the incident occurred. This is conveyed by visually indicating which sensors that activated the braking action and also by giving a short explanation as to why the braking occurred. This is done so that trust does not get so severely compromised and also so that the driver will understand that the system works properly even if an incident occurs.

Figure 62 - Giving error information in case of an incident.

9 Discussion and Recommendations

This chapter presents an overview of the thesis and its contributions, a discussion of the literature study, user study, framework, validation test, and final example HMI concept, and concludes by describing future work and recommendations.

9.1 Overview and Contributions

The project goal was how to create an appropriate level of user trust for future autonomous driving (AD) vehicles. Accomplishing this involved first identifying the factors affecting trust within the human-machine interaction in autonomous vehicles. These factors were then placed into a trust-based framework formulated around a driving scenario. The framework is intended to serve as a guide for designing trust into human-machine interaction systems in AD vehicles. The main research questions posed were:

What are the factors affecting trust in a human-machine interaction, specifically in the context of autonomous driving vehicle systems?

What events during the human-machine interaction in an autonomous driving vehicle are affected by these factors?

How can a trust-based framework contribute to designing trust into HMI systems in AD vehicles?

The questions were addressed in the following way. First, trust-affecting factors were identified in a literature study and corroborated in a user study. Second, these factors were then placed into a framework formulated around a driving scenario. The scenario was broken down into different events in order to more specifically target the appropriate placement of the trust factors within an interaction between the driver and an autonomous vehicle. Third, the framework was utilized to develop an example HMI concept.

The contributions of this thesis include the developed trust-based framework, which is intended to guide HMI designers when designing autonomous driving systems so that their customers trust these systems. This framework has gathered together direct and indirect trust factors, something that has not been done before. The literature often discusses one trust factor in isolation without addressing the bigger picture, which is very important because trust is affected by many factors. The framework has mapped the factors into a more holistic, comprehensive and illustrative scenario, which facilitates understanding of what is needed so that the future driver will have an appropriate level of trust for autonomous driving systems, thus promoting the correct usage of such systems. Furthermore, as simulation studies do not inherently entail risk, which highly affects trust, the validation test method illustrates the possibilities for a fairly simple yet more realistic way of testing HMI concepts.

9.2 Discussion

Literature Study

A thorough literature study covered topics from trust to HMI and automation. The literature study was conducted using Grounded Theory (Wolfswinkel, et al., 2013), using articles' citations to identify additional potentially relevant articles. Via this process, it was possible to work one's way from trust between humans to trust between humans and autonomous vehicles. This process also made it easier to identify relevant articles, which aided in identifying the main factors affecting trust. The user study was used to corroborate the factors identified in the literature; furthermore, no new trust-affecting factors arose during the user study, providing an indication that the most important trust factors had been found in the literature.

Some of the identified factors are relatively hard to define. To take an example, the trust factor 'feedback' can be interpreted and presented in a multitude of ways in an HMI system in an autonomous vehicle. As a result, it is therefore hard to exactly pinpoint the "best" type and timing of feedback, although much information regarding type and timing was gathered during the user study. The literature states that feedback is needed but not always how to best implement it. The fact that there is no "best" answer is something that continually needs to be considered.

There is also an issue of how the different trust-affecting factors interact to optimally create trust for the driver. It is not a matter of how each and every factor affects trust all by itself, but rather the combination of factors, which must be considered in order to create a holistic perspective over trust. Such a holistic perspective was also one of the main ideas behind this thesis, which the framework attempts to illustrate.

User Study

In addition to corroborating factors identified in the literature study, the user study (with a level 2 autonomous vehicle) also helped to pinpoint where the factors could best be implemented in the interaction between the driver and the AD system (and placed in the framework). The user study also helped the project owners to more fully grasp how these factors affect the driver in a real life scenario, been introduced to the issues first hand.

The results of the user study regarding trust and its related factors largely mirrored the results of the literature study. Although a level 2 vehicle was used, the issues surrounding a level 2 system and a level 3 system should be similar in terms of the usability of the system and how the driver understands it and its functions. The user study also helped better determine the different events in the driving scenario – e.g. manual mode, control transition one, "assisted driving" (not full AD) and control transition two – which will also exist in a level 3 vehicle.

Framework

The framework, based on the results of the literature and user studies, consists of different trust-affecting factors that have been matched to different events in a driving scenario. Basing the framework on a driving scenario was deemed a favorable way to present what takes place during the driver's interaction with the HMI in an autonomous vehicle. A scenario also makes it easy to understand how the chain of events occurs as well as how the trust factors change, both during a single interaction as well as in long-term interaction. The framework is created to help develop a greater understanding of how trust is built as well as how it changes with usage. Even though the scope of this thesis did not allow for the validation of all the factors connected to the different events in the framework, the development the scenario-based framework provides a good foundation for further development and validation, as well as for use as a design tool for HMI designers to build trust into the system.

Regarding the levels of the framework, "what is needed?" (on the part of the user) and "what affects trust?" were deliberately kept separate. This is because there are certain things that are crucial for even being able to perform the event in a favorable way, such as the ability to understand certain information. The trust factors, on the other hand, are not critical to being able to perform the task, although they are very influential in making people want to use the system or not.

Some trust factors are more generally applicable to most events, while other factors are more relevant for a few specific events. An example of a more generally applicable factor is 'anthropomorphism' that can and probably should be implemented in most events in the framework. This is because anthropomorphism can be a way to facilitate the communication of other feedback or information, for example why and how information can be conveyed in an anthropomorphic way. An example of a factor more specifically bound to a certain part of the scenario is 'uncertainty information', which is only needed in AD mode.

Furthermore, some trust factors change character in different parts of the scenario such as 'why and how information', where under normal circumstance only 'why' needs to be presented (since both why and how can cause user anxiety in the long run), but in safety critical situations both why and how information is needed for the best result.

The explanations of the factors in the different events are deliberately kept more general than specific so as to allow the designer to stay very open-minded and creative and not get locked into exact how something "should" be done (as there can be many "good" solutions). It is up to the designer to use the framework as a guideline rather than a rule. For example, the designers still need to test different types of feedback in a given event to see which type works the best for the concept under development. It is also necessary to take into account how the designs of the different sub-concepts interact with each other, something that is impossible to know beforehand. Another reason that the explanations are more general is that it remains to test and validate each factor in each event, including testing the possible levels of the factors in order to identify the preferable levels. This additional testing and validation is a point for future work.

Validation Test

The validation test considered the factor 'feedback' during the event 'AD mode' and tested three HMI concepts to see which level of feedback (i.e. how detailed or transparent) users favored. Results indicate that presenting feedback to increase system transparency via object recognition can increase the level of trust for the system, which is also supported by the literature. The results, however, do not clearly identify which of the three levels of object recognition would be most preferable, although it is most likely one of the two highest levels of transparency (based on the questionnaire and interview results). This is something that requires further testing, e.g. with a larger and more representative sample of participants.

In the test runs, the OR-concepts only identified predefined objects – cars and cones – at preselected places. When planning the test, it was debated whether to include intentional OR errors to see how participants reacted to obvious errors. However, this was not implemented due to the possibility of objects not being recognized by the different concepts, i.e. unintentional OR errors, which would introduce a potentially uncontrolled variable. It was therefore judged that unintentional errors together with intentional errors would be too much. Despite this, the question of how an unintentional error would affect participants' perceptions of the different OR concepts was partially addressed in the interviews, where participants indicated that the God-View Concept (with the most system transparency) would be most negatively affected if an error occurred precisely due to its higher level of system transparency.

Example Concept

The final, example HMI concept embodies solutions for most of the factors in most of the events in the framework. However, it should not be considered a finished concept proposal, but rather an example of how a concept could look when designed with the help of the trust-based framework; this in order to give ideas of how different trust factors could be incorporated. The HMI concept theoretically solves the trust issues that are presented in the framework, but further work is needed in order to know how well the different parts of the concept work together and if a proper level of trust is achieved overall.

9.3 Future Work

During the user study, a level 3 autonomous vehicle was not available for testing, which would likely have helped more fully grasp the true reality of the trust-related issues with these types of vehicles. Instead, the available level 2 vehicle was used (Mercedes E-class with Distronic Plus with steering assist) and the participants had to imagine that it was more self-driving than it really was. As a consequence, the focus lay more on the usability of the HMI system, something that also affects trust. However, since certain actions and events in the test vehicle were similar to those in a level three vehicle – activating, deactivating, understanding the functions, and so on – the data gathered were still relevant for study and analysis. Ideally, future user studies on level 3 systems should take place in a level 3 vehicle.

This issue of using a right-hand driven, level 2 vehicle also potentially affected the validation test of the OR-concepts. The post-test interviews revealed that the participants' beliefs regarding the vehicle's level of autonomy during the test spanned a wide range, from fully autonomous to semi-autonomous. Although this did not likely affect the test outcome, as the purpose was to create uncertainties in the autonomous driving experience (since risk and trust are interrelated), ideally the participants should have a more cohesive understanding of the level of autonomy, so as to create a more controlled experiment. Also, as discussed previously, that the test should actually take place in a level 3 vehicle.

The quality of the test conditions could also be improved to make the test more realistic, especially the timing of the animations and the fact that the concept only showed predetermined objects. Since the concepts were shown on a tablet connected to the controlling computer via WiFi, there were a few delays in the animations leading to problems with the timing in the test. The fact that only predetermined objects showed up in the concept and no other natural objects, such as trees, was something that was commented on during the interviews. To make the test more realistic and the results even more valid, these aspects should be improved upon in future work.

Another aspect of the user study, but particularly of the validation test, to be improved upon in future work is the participant sample size. Furthermore, there were issues of representativeness, e.g. the participants all had an engineering background and, for legal purposes, were all affiliated with Volvo Car Corporation, which could create bias. The participants were also in the same age group, 20 to 30 years of age, except for one individual who was over 50. In future work, it would be preferable to have a larger, more representative sample for generalization purposes.

The framework's events and factors need to be further validated by testing different HMI concepts in order to see how different levels of certain factors affect trust (e.g. levels of transparency, levels of detail of information). This thesis has only tested object recognition with different levels of transparency for the trust factor 'feedback' during the event 'AD mode'. In order to more thoroughly assure the framework's relevance, further thorough testing of the factors and levels is necessary.

9.4 Recommendations

The biggest insight from this work has been that it is not only a matter of understanding how each particular factor affects trust all by itself, but it is also a matter of considering the *interaction of trust factors* in designing HMI that creates a proper level of trust. We would also argue that it is difficult to look at just one, isolated event since *trust formation is a dynamic process* that starts long before first contact with the system. It is important to consider the whole chain of events as one unit, already starting with implicit information (e.g. commercials). Furthermore, since the *trust factors change* both during one interaction (in different events) as well as over usage, it is important to develop *an HMI concept that adapts* to the different trust factors in the various events, rather than having a static HMI concept. If one were to summarize the authors' recommendations in one sentence, it would be for HMI designers and automotive vehicle manufacturers to take *a more holistic perspective on trust*.

Bibliography

Abbink, D. A., Mulder, M. & Boer, E. R., 2012. Haptic shared control: smoothly shifting control authority?. *Cogn tech work,* Volume 14, pp. 19-28.

Anon., n.d. *Who we are and what we do; NHTSA.* [Online]
Available at: http://www.nhtsa.gov/About+NHTSA/Who+We+Are+and+What+We+Do

Barber, B., 1983. *The Logic and Limits of Trust.* New Brunswick: Rutgers University Press.

BBC, 2014. *BBC News.* [Online]
Available at: http://www.bbc.com/news/technology-25653253
[Accessed 06 May 2015].

Beller, J., Heesen, M. & Vollrath, M., 2013. Improving the Driver-Automation Interaction: An approach Using Automation Uncertainty. *Human Factors,* 55(6), pp. 1130-1141.

Bergman, B. & Klefsjö, B., 2010. *Quality - From Customer Needs to Customer Satisfaction.* 3:6 ed. Lund: Studentlitteratur AB.

Chalmers University, 2005. *MDI/Interaktionsdesign, Interaktiondesign - Analysmetoder 5p.* [Online]
Available at:
http://www.cse.chalmers.se/research/group/idc/ituniv/kurser/06/analys/metodappendix.pdf
[Accessed 04 May 2015].

Continental, 2015. *Continental - The Future in Motion.* [Online]
Available at: https://www.conti-online.com/www/automotive_de_en/themes/passenger_cars/ov_automated_driving_en/a_highway_chauffeur_en.html
[Accessed 12 May 2015].

Davidsson, S. & Alm, H., 2009. *Applying the "Team Player" Approach on Car Design,* Berlin Heidelberg: Springer-Verlag.

Dekker, S. & Woods, D., 2002. MABA-MABA or Abracadabdra? Progress on Human-Automation Co-ordination. *Cognition, Technology & Work,* Volume 4, pp. 240-244.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A. & Pierce, L. G. B. H. P., 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies,* Issue 58, pp. 697-718.

Endsley, M. R., 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors ,* 37(1), pp. 32-64.

Eriksson, M. & Lilliesköld, J., 2004. *Handbok för mindre projekt.* 1 (4) ed. Stockholm: Liber AB.

Fishmana, R. & Khanna, T., 1998. Is trust a historical residue? Information flows and trust levels. *Journal of Economic Behaviour & Organization ,* Volume 38, pp. 79-92.

Flemisch, F. et al., 2008. *Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation.,* Maastricht, Netherlands: Shaker Publishing.

Frost & Sullivan, 2015. Automotive and Transportation. *Technical Insights*, 2 January, pp. 1-12 .

Gasser, T. M. et al., 2013. *Legal consequences of an increase in vehicle automation,* Bergisch Gladbach: BASt (Federal Highway Research Institute).

Hancock, P. et al., 2013. Human-Automation Interaction Research: Past, Present, and Future.. *Ergonomics in Design: The Quarterly of Human Factors Applications,,* 21(2), pp. 9-14.

Helldin, T., Falkman, G., Riveiro, M. & Davidsson, S., 2013. *Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving.* Eindhoven, ACM.

Heslin, P. A., 2009. Better than brainstorming? Potential contextual boundary conditions to brainwriting for idea generation in organizations. *Journal of Occupational and Organizational Psychology,* Volume 82, pp. 129-145.

Hoc, J.-M., 2000. From human - machine interaction to human - machine cooperation. *Ergonomics,* 43(7), pp. 833-843.

Hoff, K. A. & Bashir, M., 2014. Trust in Automation: integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* pp. 1-28.

Jenkins, M. P., Wollocko, A., Farry, M. & Voshell, M., 2015. *Low-Level Automation as a Pathway to Appropriate Trust in the PED Enterprise: Design of a Collaborative Work Environment,* s.l.: International Command and Control Research Technology Symposium.

Jian, J.-Y., Bisantz, A. M., Drury, C. G. & Llinas, J., 1998. *Foundations for an Empirically Determined Scale of Trust in Automated Systems,* Buffalo: Center for Multisource Information Fusion Department of Industrial Engineering State University of New York at Buffalo.

Johannesson, H., Persson, J.-G. & Pettersson, D., 2004. *Produktutveckling - effektiva metoder för konstruktion och design.* 1 ed. Stockholm: Liber AB.

Johannesson, H., Persson, J.-G. & Pettersson, D., 2013. *Produktutvekling - Effektiva metoder för konstruktion och design.* s.l.:Liber.

Koo, J. et al., 2014. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance.. *International Journal on Interactive Design and Manufacturing (IJIDeM).*

KPMG, 2013. *Self-Driving Cars: Are We Ready?.* [Online].

Lee, J. D. & See, K. A., 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors,* 46(1), pp. 50-80.

Leohold, P. D. J., 2011. *Highly Automated Driving: Fiction or Future?.* s.l., HAVEit , p. 13.

Marinik, A. et al., 2014. *Human Factors Evaluation of Level 2 And Level 3 Automated Driving Concepts: Concepts of Operation ,* Washington, DC: U.S Department of Transportation, National Highway Traffic Safety Administration .

Marsh, S. & Dibben, M. R., 2003. The role of trust in information science and technology.. *Annual Review of Information Science and Technology,* Volume 37, pp. 465-498.

Mayer, R. C., Davis, J. H. & Schoorman, F. D., 1995. An integrative model of organizational trust. *Academy of Management Review,* 20(3), pp. 709-734.

McKenzie, S., 2004. *Social Sustainablity: Towards Some Definitions ,* Magill, Australia : Hawke Research Institute .

Mcknight, D. H. & Chervany, N. L., 2000. *What is Trust? A Conceptual Analysis and an Interdisciplinary Model.* s.l., AIS Electronic Library, pp. 1-8.

Merat, N., Jamson, A., Lai, F. & Carsten, O., 2012. Highly Automated Driving, Secondary Task Performance, and Driver State.. *The Journal of the Human Factors and Ergonomics Society,* 54(5), pp. 762-771.

Merat, N. & Jamson, H. A., 2009. *How do drivers behave in a highly automated car?.* s.l., s.n.

Mercedes-Benz, 2015. *Mercedes-Benz Distronic Plus with Steering Assist.* [Online]
Available at: http://techcenter.mercedes-benz.com/en/distronic_plus_steering_assist/detail.html
[Accessed 13 May 2015].

Mercedes-Benz, 2015. *Techcenter.* [Online]
Available at: http://techcenter.mercedes-benz.com/en/distronic_plus_steering_assist/detail.html?
[Accessed 21 04 2015].

Merrit, S. M. & Ilgen, D. R., 2008. Not all trust is created equal: dispositional trust and history-based trust in human-automation interaction. *Human factors,* 50(2), pp. 194-210.

Michanek, J., Breiler, A. & Books, B., 2007. *Idéagenten 2.0 en handbok i idea management.* Malmö: Bookhouse Publishing AB.

Muir, B. M., 1987. Trust between humans and machines, and the design of decision aids. *International Journal Man-Machine Studies,* Volume 27, pp. 527-539.

Muir, B. M., 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics,* 37(11), pp. 1905-1922.

Muir, B. M. & Moray, N., 1996. Trust in automation. Part II. Experimental studis of trust and human intervention in a process control simulation. *Ergonomics,* 39(3), pp. 429-460.

NHTSA, 2013. *U.S. Department of Transportation Releases Policy on Automated Vehicle Development.* [Online]
Available at:
http://www.nhtsa.gov/About+NHTSA/Press+Releases/U.S.+Department+of+Transportation+Releases+Policy+on+Automated+Vehicle+Development
[Accessed 12 May 2015].

Nielsen, J., 1993. *Articles.* [Online]
Available at: www.nngroup.com/articles/iterative-design

NVIDIA, 2015. *NVIDIA Drive PX.* [Online]
Available at: http://www.nvidia.com/object/drive-px.html
[Accessed 12 May 2015].

Parasuraman, R. & Manzey, D. H., 2010. Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors,* 52(3), pp. 381-410.

Parasuraman, R., Sheridan, T. & Wickens, C., 2008. Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs.. *Journal of Cognitive Engineering and Decision Making,* 2(2), pp. 140-160.

Pritchard, J., 2015. *Google acknowledges 11 accidents with its self-driving cars: Big Story AP.* [Online]
Available at: http://bigstory.ap.org/article/297ef1bfb75847de95d856fb08dc0687/ap-exclusive-self-driving-cars-getting-dinged-california
[Accessed 25 May 2015].

Saffarian, M., De Winter, J. & Happee, R., 2012. *Automated Driving: Human-factors issues and design solutions,* s.l.: HUMAN FACTORS and ERGONOMICS SOCIETY 56th ANNUAL MEETING .

Shaughnessy, J., Zechmeister, E. & Zechmeister, J., 2000. *Research Methods In Psychology.* 5:th ed. s.l.:McGraw-Hill.

Stanton, N. & Young, M., 2000. A proposed psychological model of driving automation. *Theoretical Issues In Ergonomical Sceince,* 1(4), pp. 315-331.

Sustainable Mobility, 2013. *Autonomous Cars.* [Online]
Available at: http://www.sustainable-mobility.org/resource-centre/month-issue/autonomous-cars.html?section=1
[Accessed 22 June 2015].

Szymaszek, A., 2014. *Between challenge and solution STATUSCOPE - a tool for designing status comunication in highly automated vehicles ,* Gothenburg, Sweden : Chalmers University of technology.

Thill, S., Nilsson, M. & Hemeren, P., 2014. *The Apparent Intelligence of a System as a Factor in SA*.

Toffetti, A. et al., 2009. *CityMobil - Human Factor Issues Regarding Highly Automated Vehicles on eLane,* s.l.: Transportation Research Record: Journal of the Transportation Research Board. .

van Schijndel-de Nooij, M. et al., 2010. *Definition of necessary vehicle and infrastructure systems,* Brussels: SMART.

Waytz, A., Heafner, J. & Epley, N., 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology 52,* pp. 113-117.

Verberne, F. M., Ham, J. M. & J.H., C., 2012. *Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars,* Eindhoven: Department of Human-Technology Interaction Eindhoven University of Technology.

Wickens, C. et al., 2010. *Stages and levels of automation: An integrated meta-analysis..* s.l., Human Factors & Ergonomics Society , pp. 389-393.

Wolfswinkel, F., Furtmueller, E. & Wilderom, P. C., 2013. Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems 22,* pp. 45-55.

Volvo Car Group, 2014. *Press Realeses.* [Online]
Available at: https://www.media.volvocars.com/global/en-gb/media/pressreleases/145619/volvo-car-groups-first-self-driving-autopilot-cars-test-on-public-roads-around-gothenburg
[Accessed 15 March 2015].

Appendices

APPENDIX I Work Breakdown Structure (WBS)

APPENDIX II Gantt-Chart

APPENDIX III User Study - Interview script

APPENDIX IV Ideation session TRUST students

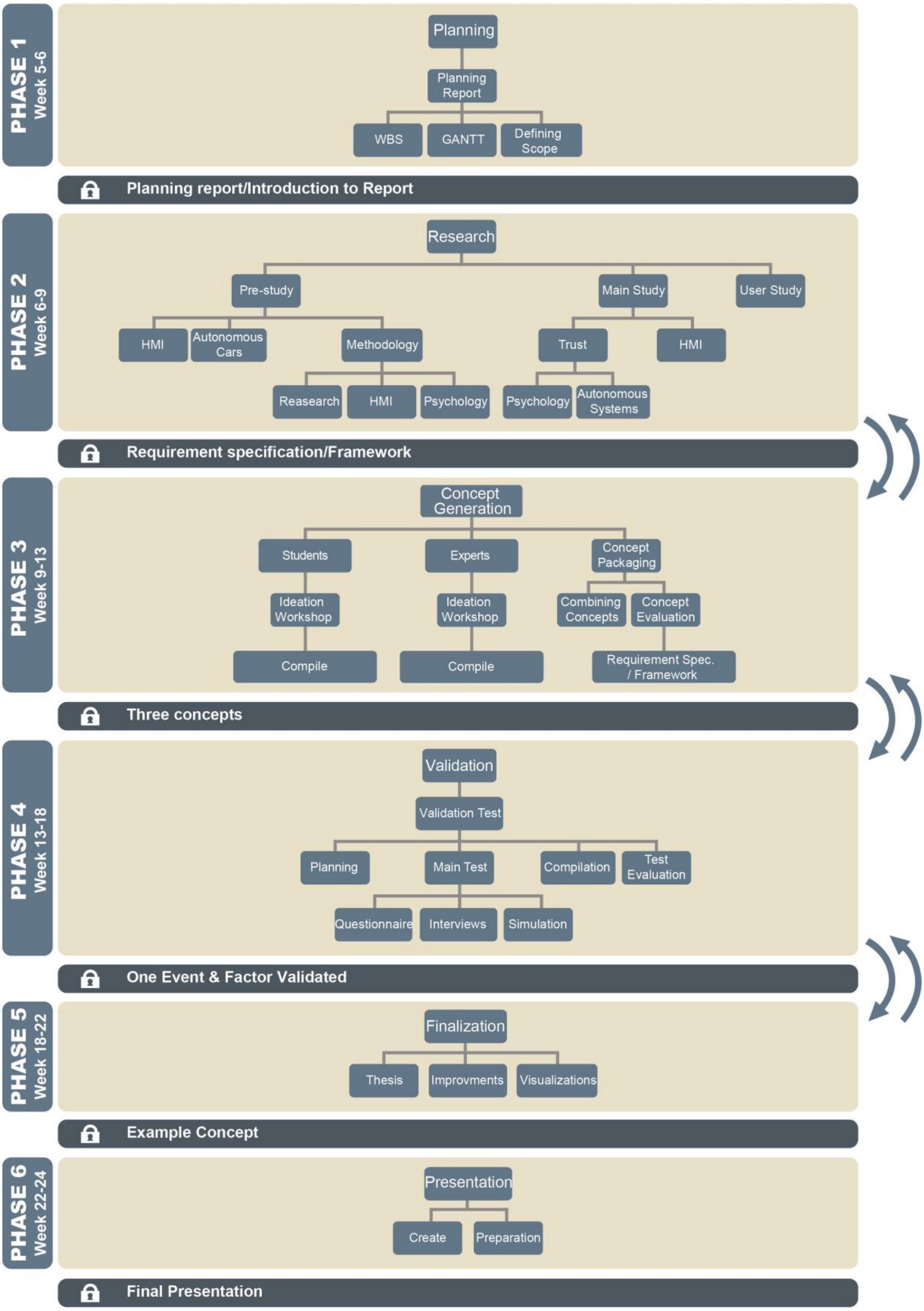
APPENDIX V Ideation session TRUST experts

APPENDIX VI Initial Analysis

APPENDIX VII Trust in Automation Questionnaire

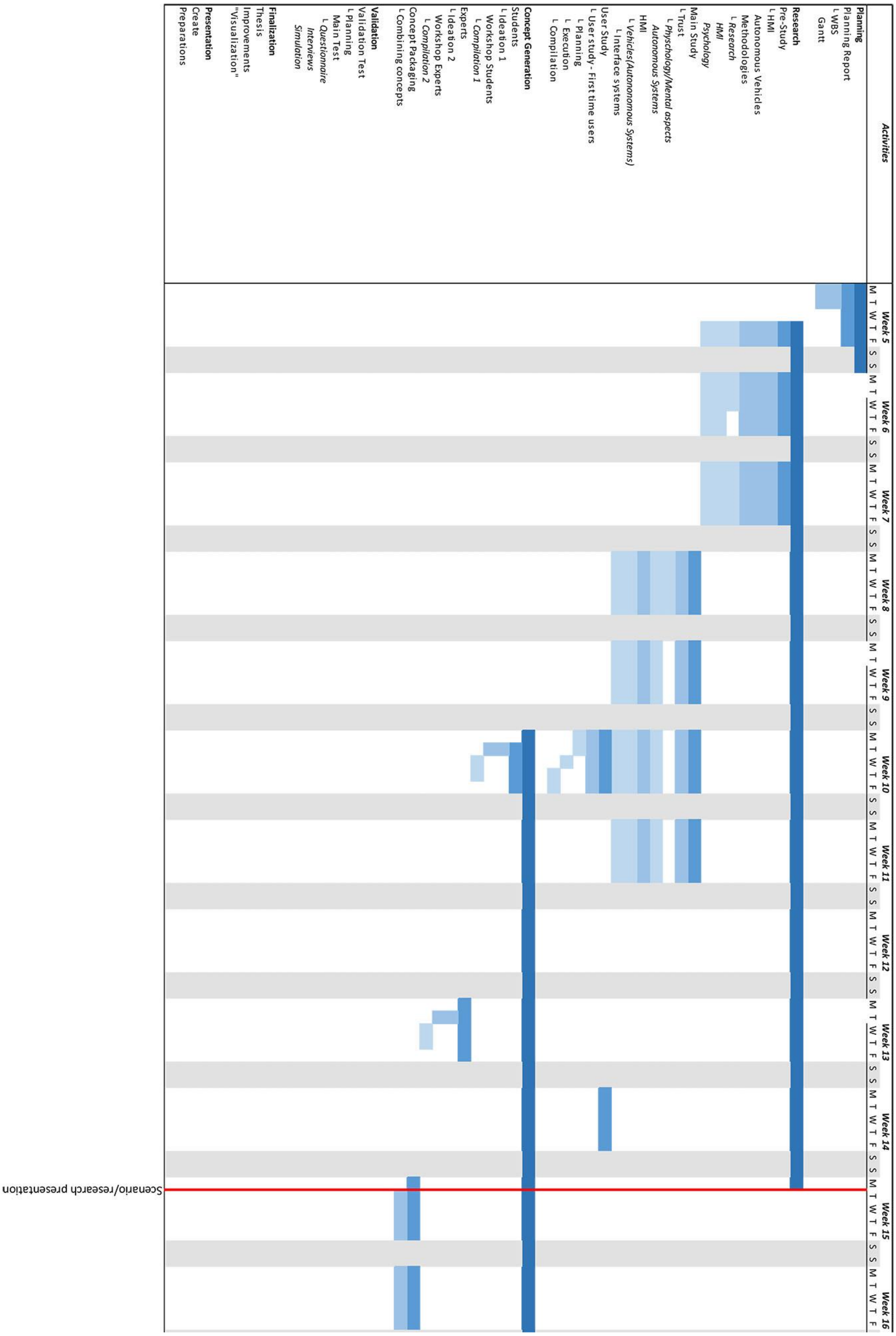
Appendix I

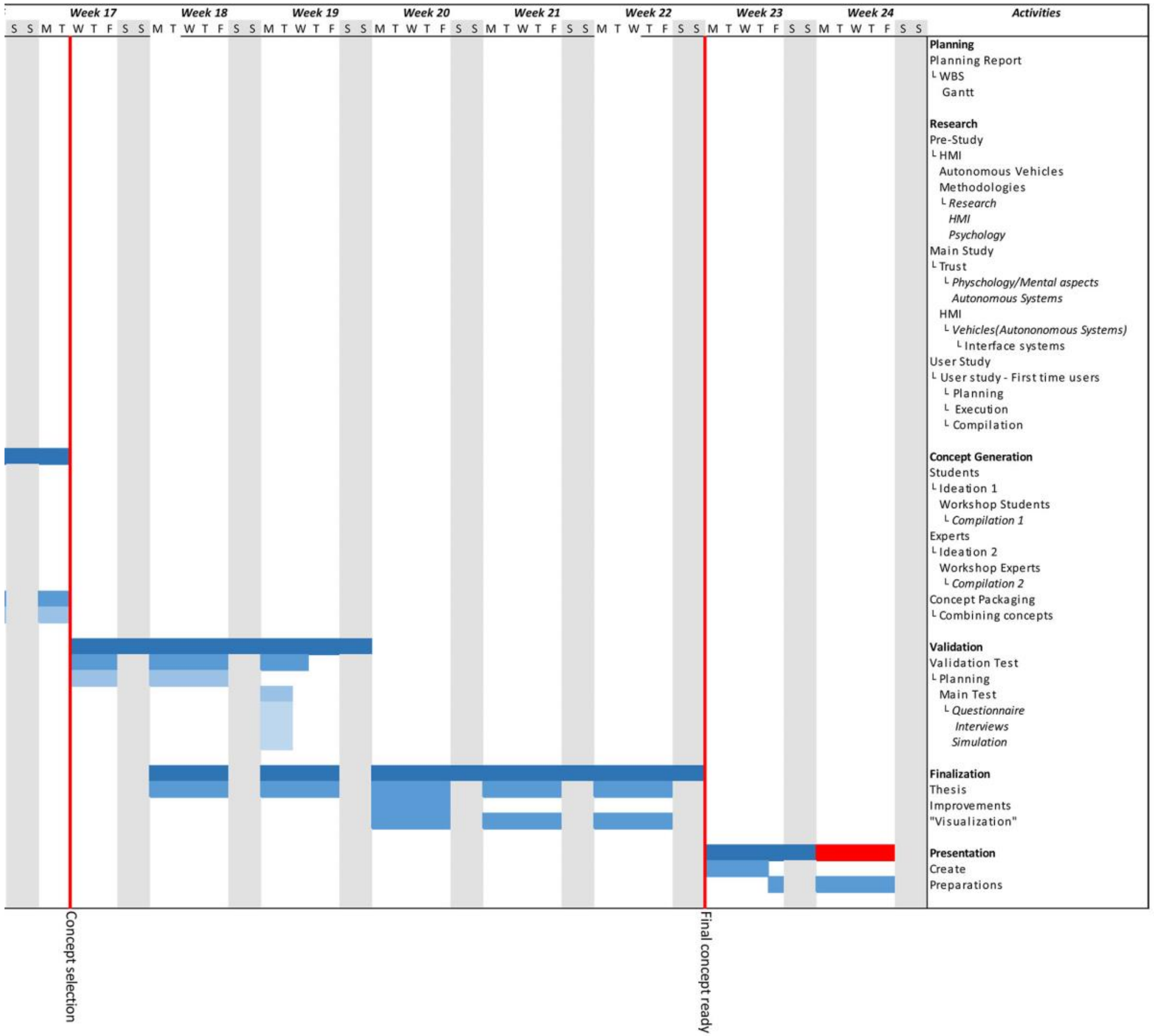
Work Breakdown Structure (WBS)



Appendix II

Gantt-Chart




Week 17
Week 18
Week 19
Week 20
Week 21
Week 22
Week 23
Week 24
Activities
Planning
Planning Report
└ WBS
Gantt
Research
Pre-Study
└ HMI
Autonomous Vehicles
Methodologies
└ Research
HMI
Psychology
Main Study
└ Trust
└ Physchology/Mental aspects
Autonomous Systems
HMI
└ Vehicles(Autononomous Systems)
└ Interface systems
User Study
└ User study - First time users
└ Planning
└ Execution
└ Compilation
Concept Generation
Students
└ Ideation 1
Workshop Students
└ Compilation 1
Experts
└ Ideation 2
Workshop Experts
└ Compilation 2
Concept Packaging
└ Combining concepts
Validation
Validation Test
└ Planning
Main Test
└ Questionnaire
Interviews
Simulation
Finalization
Thesis
Improvements
"Visualization"
Presentation
Create
Preparations
Concept selection
Final concept ready

Appendix III

User study

Interview script

How many years have you been driving?
How often do you drive? (days per month on average)

<u>Manual driving (before activation)</u>

- What do you feel?
  - The system is deceptive?
  - I am suspicious about the systems intent, action or outputs?
  - The systems actions will have a harmful or injurious outcome?
  - I am confident in the system?
  - The system provides security?
  - The system is dependable?
  - I can trust the system?
  - I am familiar with the system?

- Do you feel stressed because of the system?
  **Scale 0-100%*
- What do you expect?
- Any information you would like to have?

<u>System Activation</u>

- Do you understand how to activate the system?

- What did you feel when activating?
  - The system is deceptive?
  - I am suspicious about the systems intent, action or outputs?
  - The systems actions will have a harmful or injurious outcome?
  - I am confident in the system?
  - The system provides security?
  - The system is dependable?
  - I can trust the system?
  - I am familiar with the system?

- Do you know which functions that are activated right now?
- Did you get enough information from the systems itself in order to understand what to do?
- What type of input did you get?
- Which info helped you to activate the system?
- What info would you have liked to have more of?
- What info would you have liked to have less of?
- Any situations where it did work as expected?
- Do you feel stressed because of the system?
  **Scale 0-100%*

Assisted Driving

- Do you understand when/when not the system will assist?
- Did/do the system work as you thought it would?
    - why/why not?
- What do you feel?
    - The system is deceptive?
    - I am suspicious about the systems intent, action or outputs?
    - The systems actions will have a harmful or injurious outcome?
    - I am confident in the system?
    - The system provides security?
    - The system is dependable?
    - I can trust the system?
    - I am familiar with the system?
- What would give you more trust for the system?
- Where do you think your focus lies, on the system itself or on the road?
- What do you do while the system is activated/or think you would do?
- Is there any information you are missing or is there information that could have been presented in any other way?

Deactivation

- Did/do you know when you are in full control again?
    - How did you know?

After

- What situations did you “doubt” the system?
- Did the system work as you expected?

Appendix IV

Ideation session TRUST students

Schedule:

Brainpool (5 min)

How do you create trust?

Presentation of the subject area… Trust

6-3-5 (30 min)

- How can you increase SA?
- How can you minimize mode confusion?
- How can you make an HMI system more user-friendly through an interface? (visual, audio or haptic)
- How can you make an HMI system adapted to fit everyone? (adaptive(auto)/customization(manual))
- How would it be possible to increase the "feeling" of teamwork with an self-driving autonomous system? (Shared goal, info)
- Which information would you like to have in an autonomous vehicle system?

Open discussion (60 min)

- Mental model
    - How does your mental model look?
    - How do you think it should be?
    - How should you present the system in order to make it look like an expert?
    - What defines an expert system according to you?
- Workload (Stress)
    - How could you lower the mental workload but at the same time be alert?
    - How could you use all the different senses in an optimal way?
- System transparency
    - What type of feedback would you like to have from the system in order to have trust/transparency?
    - Which type of information do you need in order to predict the car's behaviors?
    - How should they be presented?
- Human-like (anthropomorphic)
    - Would you trust a system that has more human features than an ordinary machine/system?
    - Human-like? To what degree?
    - How would it be manifested?
- Two way communication
    - How would a two-way communication between a user and an autonomous system look according to you?
    - What type of info would you like to give and receive? And how would it be manifested?

Further opinions? Questions? Ideas?

Thank you!

Appendix V

Ideation session TRUST experts

Schedule:

Brainpool (5 min)

How do you create trust?

4-3-5 (20 min)

- How can you increase SA?
- How can you make an HMI system adapted for everyone?
- Which information is needed for the driver in an autonomous vehicle system?

2-2-3 Group Method (128 min)

- Mental model (12 min)
  - How should you present the system in order to make it look like an expert?
  - How can you create a correct mental model before first use?

Concept creation (10 min)

Presentation & Discussion (10 min)

- Workload (Stress) (12 min)
  - How could you minimize stress but still keep the driver alert?
  - How would you like to use your different senses (haptic, visual, auditory) in order to receive system information?

Concept creation (10 min)

Presentation & Discussion (10 min)

- Human-like (anthropomorphic) (12 min)
  - How would an anthropomorphic system look?
  - What kind of features would make you trust another person?

Concept creation (10 min)

Presentation & Discussion (10 min)

Thank you!

Appendix VI

Initial Analysis

Anthropomorphism

Lead words
- Playful
- Respect
- Silence
- Calm
- Kind

Physical design ques
- Smiley that shows mood
- Exterior anthropomorphism
- Virtual character that you can talk to

Characteristics
- Adjusts language to user
- A common language
- Voice & Speech
  - Saying unnecessary things i.e. nice weather
  - Talk to co-passengers as well
  - Like talking to a friend
  - Only relevant info.
  - Compliments
  - Kind respond
  - Yells at you
  - Different voices for different modes
- Natural language
- Have a name

Expert

Mediated
- Show that it works all time
- Information about proven expertis
- High end technical system

Actual
- Don´t leave anything for interpretation
- Fail-proof
- Exceeds expectations
- Does what it says
- troubleshooting & optimizing at start

Adaption

Over Time
- Plays music according to situation
- Less information after time
- Adapt to different usage levels

Context Specific
- System recognise drivers state
- System reacts to drivers actions, could suggest improvements
- Adapts to driver daily patterns

User Specific
- Knows you better than you know yourself
- Personalization of Interior
- Different level of info depending on driver
- Change interface depending on Driver
- Safe user profles in Cloud

Feedback Placement

Locating
- Use sound to lead driver to info
- Using light to locate

Accessability
- Information is where you are looking
  - If looking back, the environment in front are projected
- Virtual reality googles
- Steering wheel indicates information
- Show environment in users device (Ipad)
- Head-Up Display
- See environment in periphery

Error Information

Explain
- At error inform why
- Inform how to correct error
- Inform when user makes error

Openness

Uncertainty
- No embellishment
- Admits mistakes
- System status
- Inform about when SA is needed/not needed
- Give feedback about system certainty

Customization

Able
- Choose what information should be shown
- Being able to decide which info that should be more prominent
- Be able to have and change other than system information

Not Able
- It´s a pre-set interface
- Only give feeling that it is personizable

Mental Model

Creating Expectations
- The system have been used and tested
- Show that the system is safe
- Make people the right kind of scared
- Show the truth about the systems capabilities
- Show how systems saves lifes

Two-Way Communication

Suggestions
- Gives suggestions about possible actions
- Non-Driving suggestions (Music)
- Car proposes different participation depending on user level(age, usage etc)

Common Goals
- Ability to choose driving style(Eco-friendly mode)
- The system gives rewards (massage)
- The system asks to be washed

Ways of Communication
- Speech Control
- Hand Gestures

Transparency

Predictability
- Non-unexpected Events
- Pre-warnings
- Continous information flow about what is happening

Distinctness
- Right action to right information
- Knowing what the system does

Show system understanding
- Let the system only provide information that is true
- Show that the car has seen the objects
  - Augmented reality warn about traffic situations
  - Inform distance to vehicle ahead
- Show cars faith in other vehicles
- System must show that it sees a possible problem as well as giving information how it will proceed

Aesthetics

What
- Show peacefull environment in screens/windows
- Pictogram
- Clear and well-known symbols

How
- Use 80´s aestetics, very technical looking
- Hierarchical information (i.e. important=big)
- Make it pretty

Gamification/Making Interesting

Fun
- After learning phase make the system more playful
- Onboard games were you can compete against the system

Promote Learning
- Show learning through having levels so that you can improve your knwoledge about the system (get stars)
- Make information interesting so that the driver wants to participate
- Make functions not scary to try

Simplicity

Easyness
- Hide too technical information
- Only simpel choices
- Less functions
- Easy, intuitive and clear interface
- As little information as possible
- No physical buttons

Safety Functions

Control
- Always be able to take control
- Deadmans Grip
- Auto-Off Function

Training

Test
- Test the system through a simulator
- Testdrive the vehicle and by that the system
- Test the worst case scenarios so that you know how to handle them

Tutorial
- Introvideo showing what affects driving
- Lectures and/or workshops with the intent to learn more about the system
- A instruction video before every usage(flight security video)
- Step-by-step tutorial before first usage
- Have a demo before using a certain function (first time)

Operator Assistance
- Outsource the AD to an external operator

Familiarity

Recognizable
- Familiar objects/interfaces
- Common ground
- Be like me

Continuity
- "Same old story"
- Experience
- Proven system

Soft Transition

Pre-information
- System gives pre-warnings a long time before control transition happen
- Visual graph shows how close transition is

Physical cues
- Make the driver want to look out the window when approaching a transition
- Seat/steering wheel changes to an upright/driving position when going from AD to manual
- Make it uncomfortable to do other things when approaching an transition
- Change color in car when an task is emminent

Assistance
- System provides extra assistance a short while after transition

Friendly

Caring
- Reassure
- Always be there
- Cares about drivers safety above all
- System corrects driver behavior in a polite way

Attentiveness
- System makes me laugh
- System prioritize´s my needs and wishes
- System aknowledges me
- Makes the driver feel like an expert
  - Superlative praising
  - Never makes me feel stupid
  - Driver should feel in control

In-the-loop

Continous Feedback
- Regular system check
- System has a continous dialog with the driver
- Vibrations in seat on a regular basis to keep driver in the loop
- Make the driver look at the road continously
- Direct the drivers attention towards important situations

Give environmental information
- Guided tour
- Always have a map available
- Inform of ongoing activities in the surroundings

Senses

Sound
- Calming music
- Sound when system is malfunctioning
- Change volume on stereo when attention is needed
- Soft sound

Visual
- Calming colors and lights
- An special AD color as well as a special manual mode color

Haptics
- Connect to a smart watch that vibrates
- The seat adapts to an upcoming action
  - Seat rest raises/lowers
  - Seat stiffens up
- Only use haptic when AD mode is active
- Use a vest that convey haptic information

Combination
- An direct brain feedback(connection between system and driver through brain sensors)
- Use a combination of all three elements; sound, visual and haptics
  - Increasing levels(from haptic to sound)
  - Feedback specific sensory output

System Understanding

Mode Clarity
- Warnings between AD-levels
- Clearly show which mode is active
  - Color in display changes
  - The physical interior changes
  - The color of the interior changes
  - Like automatic gearboxes,show which gear is active
- Few and clear AD levels
- Clearly show if AD is on and who is control

In-Depth Information
- If driver needs more information it should be possible to get
- Show it through visualizations
- Understand limitations
- Understand the holistic perspectiv
- Make it possible to see where the systems gets it´s information from

Why Information
- Wants to know why actions are taken

Summary
- Summary after drive about events and actions taken

Quality

Competence
- Fail proof
- Should always work
- A natural movement pattern (soft)
- Quality feel

Appendix VII

Trust in Automation Questionnaire

OR-Concept XX

Participant no: XX

Q1: I understand how the system works – its goals, actions and output

1 2 3 4 5 6 7

Q2: I would like to use the system if it was available in my own car

1 2 3 4 5 6 7

Q3: I think that the actions of the system will have a positive effect on my own driving

1 2 3 4 5 6 7

Q4: I put my faith in the system

1 2 3 4 5 6 7

Q5: I think that the system provides safety during driving

1 2 3 4 5 6 7

Q6: I think that the system is reliable

1 2 3 4 5 6 7

Q7: I can trust the system

1 2 3 4 5 6 7

CHALMERS UNIVERSITY OF TECHNOLOGY
SE 412 96 Gothenburg, Sweden

Phone: + 46 (0)31 772 10 00
Web: www.tekniskdesign.se