# Safe Transitions Between a Driver and an Automated Driving System

Rolf Johansson
Zenuity, RISE
Göteborg, Sweden
e-mail: rolf.johansson@zenuity.com

Jonas Nilsson
Zenuity
Göteborg, Sweden
e-mail: jonas.nilsson@zenuity.com

Annika Larsson
Autoliv Development
Vårgårda, Sweden
e-mail: annika.larsson@autoliv.com

***Abstract*—This paper presents a methodology for achieving functional safety for an automated driving system (SAE Level 4) with respect to safe transitions between the driver and the system. Safety analysis and assessment of an implementation example show how to allocate safety requirements on Human-Machine Interface (HMI) components to handle the risks of unfair transition, mode confusion and stuck in transition, respectively. The methodology is appropriate for different assumptions on driver failures. The paper shows how to identify safety requirements on the HMI components, given that there is an assumption of a set of single, double or multiple failures by the driver. Results from this example show that it is sufficient to allocate safety requirements on the sensor and the lock of a control to ensure safe transitions. No safety requirements are needed on visual feedback to the driver, e.g., displays.**

***Keywords-functional safety; automated driving system; HMI; safety assessment.***

## I. INTRODUCTION

Presently, the most critical factor for road vehicle safety is the behaviour of the driver. There are different estimates, but a common understanding is that driver mistakes in the last seconds before a critical situation is a contributing factor of more than 90% of serious accidents. However, drivers are competent in general to drive safely and handle most risky situations well.

The potential safety benefit of increased vehicle automation is undoubtedly large but it is important that the extra risks coming from potential failures of automation are limited to a minimum. More advanced functionality and intelligence implemented in the vehicle means that more of the responsibility to drive safely shifts from the driver to functionality implemented in the vehicle. In the discipline of functional safety, there are methods to assess risks of malfunctioning electrical/electronic (E/E) implemented functionality, and to reduce these sufficiently. For road vehicles, ISO26262 is the functional safety standard.

This paper is an extension of [1] and focuses on higher levels of driving automation in on-road vehicles, where the Automated Driving System (ADS) may be given full responsibility for a safe behaviour in traffic. According to the taxonomy and definitions in J3016 [2], we can say that on level 4 (L4) automation, the ADS inside its operating driving domain (ODD) takes full responsibility, including fallback, for the Dynamic Driving Task (DDT).

Regarding the responsibility of the driver, the precise L4 definition says that when in charge, the ADS is responsible for the DDT "…without any expectation that a *user* will respond to a *request to intervene*".

The introduction of an ADS with full responsibility for the DDT, implies that the problem of traffic safety for an L4-equipped vehicle can be decomposed into three subproblems:

1. Safe driving when the ADS is in charge
2. Safe driving when the driver is in charge
3. Safe transitions between the driver and the ADS.

The first point is obvious when stating that the ADS is responsible for driving, and may be the major functional safety challenge for L4-equipped vehicles. The second point includes specific topics coming from the introduction of L4-equipped vehicles. This is because the introduction of an ADS may lead to, e.g., that the driver by mistake relies on an inactive ADS. A special case under the second point, is when the driver makes a mistake of what vehicle he/she is inside now. Such a mistake may hence cause a traditional non-ADS-equipped vehicle to become unsafe because the driver thinks there is an ADS in it. That case is not elaborated in this paper having focus on L4-equipped vehicles. The third point is the focus of this paper.

Having functionally safe transitions, implies showing absence of malfunctions in the ADS transition functionality that may lead to unacceptable risks. This includes risks related to the interaction between the driver and the ADS. One key question is what are reasonable human mistakes and misuse? Consequently, human factors expertise is a vital part to achieve functional safety.

Note that this paper focuses on the functional safety of an ADS and as a part of this analysis use human factors expertise, not the other way around. The goal of functional safety is to show that the remaining risk for the system (e.g., the ADS) in its context is *reasonable*. Human factors (HF) in safety, on the other hand, focus on *optimizing* the safety of the driver-vehicle system, but do not have the ambition to show that all risks (due to, e.g., system malfunctions) have been sufficiently mitigated.

The contribution of this paper is a method for achieving functional safety for an L4-equipped vehicle with respect to safe transitions between the driver and the ADS. Note that according to the definition of L4, we need to achieve this under no assumptions that the driver will take back control within a bounded time.

This paper is organized as follows. Section II refers to related work. Section III describes the new hazards related to the driving mode transitions introduced by SAE L4. In Section IV, we discuss how to define a safe transition and the acceptable level of tolerance to driver mistakes. Section V elaborates on possible implementations using a system example and corresponding functional safety analysis and assessment. Finally, Section VI presents concluding remarks.

## II. RELATED WORK

As the existing autonomous systems within the automotive industry are still in their infant stages and the majority of them still are semi-autonomous (i.e., SAE L1-L2) at time of writing, these systems are excluded from the state-of-the-art comparison. The interested reader may study results from several research efforts on this topic; PReVENT, HAVE IT, ADAPTIVE and INTERACTIVE to mention a few.

### A. Related Work in Automotive

There are two general strategies how to consider the interaction between the driver and the ADS. SAE L4, which is the focus of this paper, uses by definition the more conservative strategy where there are no assumptions that the driver can take back control within a bounded time. We can call this an autopilot with full responsibility for safety, as it does not need to rely on any responsiveness from the manual driver to stay safe.

Automotive research within human factors and transitions of control have focused on the less conservative strategy and consequently SAE L2 and L3 vehicles. This is a research question that is currently very much investigated [3] [4] [5]. A rather recent overview of what controllability assumptions that are reasonable for different levels of vehicle automation is also found in [6]. Research on L3 vehicles has focused on how transitions may take shorter or longer time to complete, as drivers who are not in control will sometimes prioritize other tasks over driving. Results indicate that the time needed for a safe and completed transition of control following an ADS request for transition, ranges from about 7 seconds to over 30 seconds [3] [7]. Thus, it may be difficult to allow drivers to be out of the driving loop and still expect them to accept and succeed in resuming control within a short time period.

It has also been shown that it may take drivers several seconds to control the vehicle manually without reduced performance, following a forced resumption of manual control [8]. This reduction in controllability are however at a very detailed level, and other research has indicated that about 7-10 seconds after being forced to resume manual control, drivers are able to avoid critical situations as well as before activating automation [3].

Much research effort has also been spent on optimizing the design of the HMI to alert drivers to the need to resume control from L2 and L3 vehicles. Results have indicated that it is preferable to show the driver what the vehicle is "aware" of so drivers can handle the situation if needed [9] [10] [11]. The known accidents that have been attributed to vehicle automation so far, have been caused by automation limitations in handling the DDT and the driver feeling so safe as to cease monitoring its limitations.

There are also simulator studies suggesting that human drivers may change their driving behaviour when taking back control from an autopilot [12]. This is not considered in this paper as we focus on functional safety rather than design of the HMI or autopilot driving behaviour.

### B. State-of-the-art comparison with other industries

This section describes technology, systems and concepts from other industries where similar problems arise caused by mode confusion and unsafe transitions. The focus has been on nuclear, avionics and rail since these industries deal with complex systems, exist in a regulated environment and all demand active users for proper operations. Experiences from other industries give valuable insight into how to design interfaces and processes that ensure safe transitions in the context of autonomous driving.

The two major players in the civilian avionics industry, Boeing and Airbus, apply different philosophies regarding automation. Boeing implements a strict assisting role for technology and automation, where the pilot always acts as the final authority. Airbus rather sees automation as a way of enhancing flight performance by assisting the responsible pilot. This subtle difference in philosophy causes different problems, where the Boeing strategy allows the pilot to perform errors that may cause accidents and the Airbus strategy may interfere and prevent the pilot from performing necessary manoeuvres needed for safety in extreme situations [13] [14]. One approach cannot, however, necessarily be said to be safer than the other as aviation accidents are very rare.

In military avionics, there is a system called Auto Ground Collision Avoidance System (Auto GCAS) that monitors the pilot's response in certain situations and if the pilot does not respond to an alarm, the system takes over and performs the necessary manoeuvre. After avoiding the threat, control is returned to the pilot. Inagaki describes this as situation-adaptive autonomy where authority over a system is transferred between human and machine agents [15]. A similar system in the automotive industry is that of forward collision warning with automated emergency braking, where the warning comes first, and if the driver does not respond to the warning, automated emergency braking intervenes.

However, the main point of reference within both civilian and military avionics is that an educated pilot is always responsible for operation of the airplane with the help of coordination and information from air traffic control, differing from the automotive situation envisioned in SAE L4-5. In aviation, there are also several protocols for transitioning control, be it between pilot and co-pilot (pilot flying and pilot not flying) or between pilot and autopilot. In emergencies, civilian pilots generally have several minutes to diagnose a problem and try different countermeasures, being able to consult each other while doing so.

Within the nuclear industry there are numerous processes for operators to monitor. This is handled with different interfaces displaying process information. One main control board represents the state of the system and operators are

specially trained on how to read it. The main control board is assured to high safety integrity and acts as the primary source of information should different sources provide inconsistent information. Nuclear operators are well educated with the processes and the system and are regularly trained in handling risky scenarios. They also use binders which contain detailed information on what procedure to follow given different error messages and states of the main control board. Some tasks that could be automated have not been, in order not to make the operators passive and complacent to changes in the system state [16].

In modern nuclear power plants, there are specific procedures ensuring correct decisions are made even in emergencies regarding the operation of the nuclear plant. Regulations state that the plants are to be designed in such a way that operators always have a 30-minute window to search for, deliberate and perform a procedure. In other words, the plant is fully autonomous for 30 minutes at a time [16]. There are also mechanisms for actions at high safety levels that require several users to acknowledge the actions independently in order to perform it. The time allowed for deliberation is, thus, much longer than in automotive or any other vehicle industry.

Studies from the rail industry have analysed operator workload and the possibilities of it causing human errors. Two main ways of managing human performance have been formulated, through either technology or human resource management. Assessment of individual possibilities to manage the required workload has been performed through psychometric testing, as well as limiting workdays and issuing regular breaks [17].

When reviewing earlier experiences from the nuclear, avionics and rail industries we make three important observations. One: In nuclear, rail, avionics and space the time available to operators are on the minute scale, sometimes tens of minutes. This means incidents in those contexts allow for perception, deliberation, and action. Automotive often operates in much shorter time scales in the realms of seconds and milliseconds, leading to much shorter response times mainly allowing for perception and action. Two: Within these industries, the technical solutions are operated by educated users, certified to use the specific equipment, and trained on a regular basis. This is not the case in automotive, where most countries only require one driving test during the entire lifetime of the driver. Three: These industries rely heavily on safety procedures, regulating what is to be done and in what order. These procedures are often written on paper and can be physically viewed in case of emergencies. These industries also often operate in controlled environments and operators handle incidents in cooperation with colleagues supporting them.

Translating information displays such as those in the nuclear industry into the automotive setting is problematic, as most of the information sources' primary purpose in cars is to enhance and ease the experience rather than to provide safety-assured information on the system state. Also, the displays in nuclear demand training for the operators and are not self-explanatory. Adaptive interface features linked to specific task requirements with consistency in interface design across different modes of system operation is recommended for the users to effectively apply mental models [18]. As the automotive setting makes it difficult to limit usage periods, the technology and interfaces must be designed to ensure safe usage under these different circumstances.

## III WHAT CAN CAUSE THE ADS-EQUIPPED ROAD VEHICLE TO BECOME UNSAFE

One interpretation of a hazard analysis & risk assessment (HA&RA) today according to ISO26262 is that the vehicle itself is considered safe, if it only puts the driver in situations that are possible to manage safely. The driver is ultimately responsible for safe driving, and the malfunctions of the vehicle should be restricted in such a way that the driver can keep the vehicle in a safe state. The explicit method for determining the requested Automotive Safety Integrity Level (ASIL), restricting a certain hypothetical vehicle failure, is to measure three factors: exposure (E), severity (S) and controllability (C). The two first factors are the traditional ones that are part of the definition of risk, i.e., a combination of probability and severity. The third factor is the one that considers that the driver may sometimes have a possibility to keep the vehicle safe, even though the ordinary (safety-related) functionality is failing.

When we shift from a situation where a manual driver has the ultimate responsibility, to highly automated driving where the manual driver and the ADS are alternating, this will have an impact on the HA&RA. So, what will become different when going to SAE L4? This new challenge has partly been addressed in [19].

As a starting point, we require the same from an ADS as from a driver. This means focusing on a safe style of driving, making the driver or ADS capable to handle also unexpected events. When programming an ADS, this is what we cover on the tactical level [2] [20] [21] [22]. The ADS should always choose to perform the manoeuvres in such a way that reasonable, but still unexpected, situations could be handled safely. For example, the decision whether to initiate an overtaking manoeuvre is on the tactical level. An optimistic decision to overtake may place the vehicle in a situation where avoiding one accident may cause another. The solution to this dilemma is of course to initiate an overtaking manoeuvre only when the entire operation is foreseen to be possible to fulfil in a safe manner.

Note the contrast to Advanced Driver Assistance Systems (ADAS), where the vehicle takes over mainly on the operational time scale, maintaining as steady-state as possible, and then assumes the manual driver to continue according to the (maybe revised) tactical plan. The ADAS functionality today does not take the ultimate responsibility to drive the vehicle safely. Firstly, it operates on the operational time scale, and does not revise tactical plans. Secondly, it only assists the driver. In SAE L4 when responsibility is transferred from the driver to the ADS, there is no longer an assistance relation. The transfer means that from then on, the ADS is fully responsible for driving the vehicle safely.

Given that the ADS can drive safely once in command, the HA&RA must also cover the transitions between the driver and the ADS. In SAE L4, these transitions introduce three new types of hazards, namely *unfair transition*, *mode confusion* and *stuck in transition*. These are described in detail in the following sections.

### A. Unfair transitions

It may be complicated for the driver to make a proper override of what is perceived as a failing or unsafe tactical decision of the ADS. This is because drivers may find different tactical solutions to a certain driving situation, and each of these may be correct. It may be hard for a driver to distinguish an unsafe tactical decision from a one that is just different from his or her own favourite pattern. Even more, it may be hard to continue to fulfil a tactical plan of another driver if the responsibility is transferred in the middle of the intended sequence. This difficulty is both for a driver to continue a plan of the ADS, and for the ADS to continue what has been initiated by the manual driver. Problems can arise in terms of non-driving task engagement, safe headways, and the knowledge of other road vehicles' positions.

If the manual driver realizes that the ADS has handed over responsibility, without the manual driver agreeing to this, this is a new risk to consider when entering SAE L4. We can say that the manual driver is put in a situation of *unfair transition*. For a driver with the same understanding of the traffic situation and the control of the vehicle, the situation may be possible and easy to handle, but an unfair transition may put the driver in a situation where continuing to drive can be difficult. For example, the driver may be engrossed in a non-driving related task and therefore take a long time to resume manual control [7].

The problem of unfair transitions may appear in both directions. It is reasonable to assume that the automated driver can drive safely as long as it can choose its own tactics. This is a far easier task than being able to understand and solve arbitrary situations.

To summarize, if the responsibility is transferred from one driver to the other, this must include a confirmation from the receiving driver. Otherwise, the transition may be regarded as unfair, and it is a non-negligible risk that the second driver is incapable of handling the situation, on both operational and tactical time scales.

### B. Mode confusion

In order to make the entire trip from start to stop safe, it is critical that the two drivers always agree which of them currently is in charge. If they misunderstand each other, there is a risk that either there are two drivers trying to control the vehicle, or there is no one taking care of the ride. Both these potential *mode confusions* need to be addressed.

If we allow both the manual driver and the automated driver to override each other, there is an obvious risk that the resulting non-harmonized commanding of the vehicle may result in dangerous situations. This is especially probable because the two drivers most likely make different tactical decisions now and then, and as consequence regard the operative command of the other as faulty. For safe driving in SAE L4, it is important to reduce the risk of this reciprocal *override*. Note that this does not necessarily exclude the opportunity to adjust operational tasks such as lane position, within bounds that the responsible (driver or ADS) agrees with.

It is perhaps even more obvious that it will become dangerous if neither the manual driver nor the automated driver regard herself or himself as the ultimately responsible. Such reciprocal *underride* is therefore obviously important to reduce properly when performing the risk assessment for driving on SAE L4.

### C. Stuck in Transition

If either the ADS or the driver is unsuccessful in executing a transition for some period of time, then this may impair the driving skills of the responsible party, thus leading to a hazard. Consider the case when the driver tries to activate, e.g., by pressing one or multiple buttons, and the ADS refuses or fails to activate itself. The driver might react to this by repeatedly pressing the buttons to activate the system. When doing so, it is a risk that the driver is *stuck in transition*, gets distracted and thus cannot drive safely.

## IV. Method for Assuring Safe Transitions

In the previous section, we listed new categories of hazards to handle related to the dual driving modes when going up in automation degree to SAE L4. In the following sections, we outline a method to handle these. In Section IV-A we discuss how to define a safe handover functionality and in Section IV-B we describe how to do Hazard Analysis and Risk Assessment, HA&RA.

### A. Principles for safe handover

Below we propose a way to define the part of a functionality (denoted *item* in ISO26262 [23] and *feature* in J3016 [2]) which transfers control of the DDT between the driver and the ADS. We remark that this section only discusses the part of the item definition which is related to transitions. A complete item/feature for an ADS will also include, e.g., how the ADS drives, i.e., performs the DDT.

We assume that both the manual driver and the ADS are capable of safe driving, as well as judging its own ability to drive safely. Being capable of safe driving also includes driving safely until a handover is completed.

The item definition seeks to be "traffic safe by definition" assuming that the ADS works as intended. This is to say that functional safety of the item/feature implies traffic safety. Consequently, only violations of the principles below, i.e., malfunctions can lead to hazards. We remark that it is possible to make a more explicit definition of the item/feature functionality, which would then require that safety of the defined functionality must be proved outside the scope of functional safety.

For a safe transition of control between manual driver and the ADS, transfer of responsibility for the DDT may only occur if the following conditions are fulfilled:

1. Driver and the ADS both accept transfer, i.e., have consensus
2. The recipient (driver or ADS) is capable to drive safely

These two points introduce a fair procedure for handover to eliminate *unfair transitions*. This means that the current responsible (driver or ADS) stays responsible until there is an agreement for a handover to a capable recipient. This also implies that both the driver and the ADS need to explicitly confirm that a transition is possible and fair to perform. Furthermore, it implies that both the driver and the ADS really are aware of what has been agreed. Thus, neither the driver nor the ADS are required to take control and thus the vehicle will be in a safe state if either of them accept to take control of the vehicle.

Note that these two principles may imply that conditions on the surrounding environment are fulfilled. Traffic situation will probably need to be "tactically simple" to hand over safely from the ADS to the driver.

The problem of *Mode confusion* can be solved by combining the safe handover procedure described above with mechanisms that handle interference from the part which is not in charge, i.e., override. To ensure safe driving between transitions, the following condition must also be fulfilled:

3. The non-responsible party (driver or ADS) must not affect vehicle motion outside the constraints set by the responsible party (ADS or driver)

This can be handled either by making the current responsible capable of ignoring the other or by avoiding interference by the non-responsible party. When the driver is responsible, we require the ADS not to interfere in such a way that the driver cannot control the motion of the vehicle. This is how ADAS are designed today.

When the ADS is responsible, the driver should then try to avoid interfering with the vehicle controls. A potential solution is not allowing the driver to have any impact on the vehicle, if not first going through a handover procedure. We then transfer part of the responsibility to the ADS by putting safety requirements on ignoring any try from the driver to control the motion of the vehicle. For means of trust and comfort, it may be advisable to allow the human operator to control, e.g., lane positioning or following distance. The range of such adjustment should then be constrained within bounds set by the ADS, similar to how adaptive cruise control contains merely a few distance settings.

To manage the *stuck-in-transition* hazard we formulate the following condition:

4. Transition sequence shall not affect the capability of the responsible party (driver or ADS) to drive safely

This will put requirements on the handover sequence to be easily managed by the driver when activating the ADS. In the other direction, the ADS must not let the deactivation sequence affect its driving.

There are many ways to define a detailed handover protocol between the driver and ADS which implement the safety principles above. For examples, see Section V.

### *B. Hazard Analysis and Risk Assessment*

A Hazard Analysis and Risk Assessment (HA&RA) is needed to identify situations and driver behaviours that could lead to hazards. The driver behaviours to be analysed must also include reasonably foreseeable driver mistakes. Already today, we have a substantial amount of serious traffic accidents caused by driver lapses. There is no reason why not to regard the driver of a highly automated vehicle as prone to mistakes in any HMI, including the one for transition of responsibility.

The granularity of this analysis is a design choice. We could make a conservative assumption that all driver mistakes are common and that they will always lead to severe hazards. This makes the HA&RA simple but will most likely lead to higher ASIL on some system components compared to a more detailed HA&RA.

All hazardous events are assigned values for exposure, severity and controllability C, which together lead to an ASIL. As an example, consider the case where the ADS is driving at high speed and a malfunction in the ADS combined with a relatively frequent driver mistake leads to unintentional deactivation of the ADS. The situation is common leading to high exposure (E4). Furthermore, we assume that the driver cannot control the vehicle at unintentional deactivation (C3) and that this would have a fatal consequence (S3). Conclusion is that the system must not deactivate at high speed due to this specific driver mistake with ASIL D.

A similar analysis can be performed for any hazard and situation, e.g., single or multiple and coordinated driver mistakes. Less probable driver mistakes will result in a lower exposure and thereby lower ASIL.

A way to argue that a transition is safe with regards to all relevant driver mistakes is to check what happens if there is either a driver mistake or an E/E failure, or combination of these. This must be checked for any state in the transition protocol. For any hazardous consequence, it must be shown that the corresponding E/E failure is prevented with an appropriate safety requirement.

Note that this method also addresses the nominal function of the protocol. If a manual failure may lead to a hazardous consequence even in a fault-free case, the protocol implementation is obviously not robust enough.

For the ADS, we assume that safety requirements are allocated to all elements critical for achieving a transition in such a way that it can be considered as fair and consistently understood by both drivers. Of course, redundancy patterns may be applied allowing the ASIL D to be decomposed onto different elements of the implementation.

## V. General Implementation Suggestion

This section provides some guidance on how to design and implement a protocol for safe handovers. To make a transition tolerant to any single manual mistake, there are a few different general ways to design the protocol. The

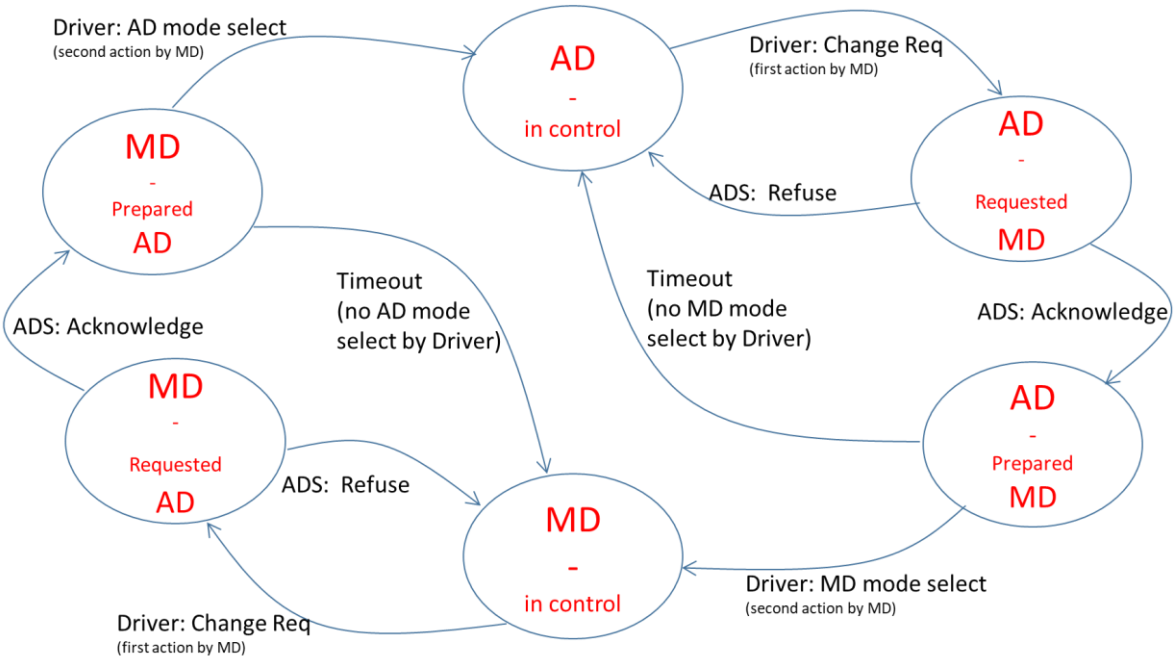


Figure 1. Example of a simple transition protocol.

redundant action from the manual driver can in general be either in time or in space, or a combination of these. By time redundancy, we mean here to request a sequence of actions where the second must follow in a certain time interval after the first one. Space redundancy is on the other hand when the manual driver is requested to apply several actions simultaneously. In both cases, the idea is that it can be argued that the set of actions is extremely unlikely to be performed by mistake.

A less conservative assumption is that a protocol should be immune against any single manual mistake. A more conservative assumption is to increase the number of mistakes a driver can perform still being conformant to the protocol. A high number of such mistakes may be argued to occur if they entire protocol sequence can be seen as an automated behaviour, being possible to execute in total by mistake. The scope of this paper is not to argue what combination of manual mistakes that are likely, but showing a general technique and illustrate this with some examples. The first examples are designed to be robust to single human mistakes, and the last one is implementing a protocol still safe in the presence of two human mistakes.

### A. *Example HMI Protocol and Implementations*

As a first example in this paper, we chose to describe a protocol based on manual time redundancy. This means that we always require two actions from the driver for any transition from the mode when the driver is responsible, here denoted MD, to the mode when the ADS is responsible, here denoted AD. The same requirement on two actions by the manual driver are also valid for the reverse transition from the mode AD to the mode MD. Furthermore, we say that the second action of the manual driver defines the transition, which means that there is no requirement on the manual driver to observe the resulting outcome correctly, more than knowing what she or he is doing herself or himself. As long as the second action is fulfilled, the transition is deemed to have occurred.

In Figure 1, a general protocol is illustrated, where two coordinated actions are required from the manual driver. When implementing this it is important for the ADS HMI not to allow the driver to perform the second action, without having acknowledged the first one.

In this example, we choose the first action to be a press of a button and the second to be a change of lever position. This lever has exactly two possible positions, equal to the two modes: AD and MD. For a certain L4 feature, the journey is always to be started in MD, and the driver may change the mode after reaching the proper state in the transition protocol. We consider the lever to be locked at any other time. Furthermore, if the lever is not moved fast enough after getting acknowledge by the ADS, it will be locked again requiring the protocol to start over again to perform a transition.

This protocol is based on the assumption that it is always safe to keep the mode if nothing else is agreed. The current responsible (manual driver or ADS) should always be able to continue to take care of the vehicle in a safe manner. The exception is when the progress of the protocol execution is hindered in a way that generates the failure *stuck in transition*.

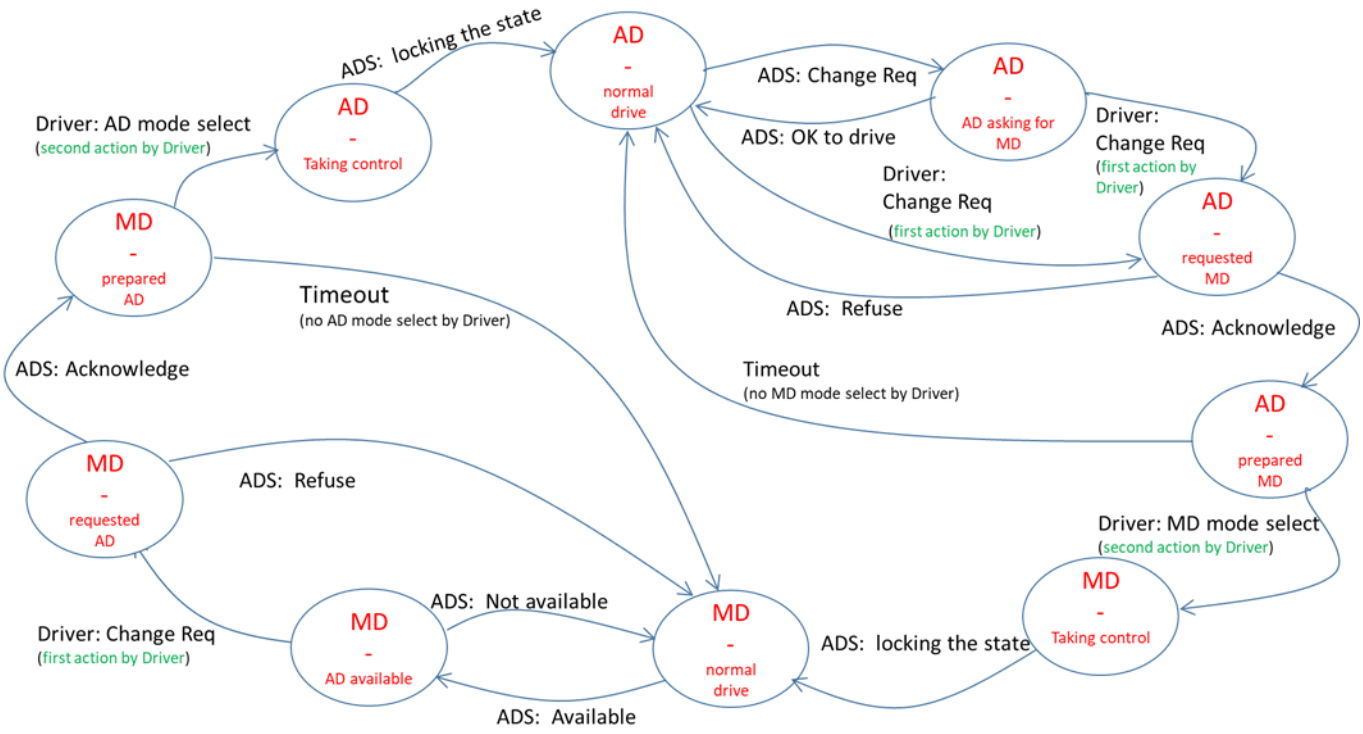


Figure 2. Example of an elaborated transition protocol.

We can extend the protocol to cover the cases where the ADS can suggest a transition, either by declaring that the ADS is ready to take over from the manual driver, or by telling the manual driver that the ADS performance is limited. Such a protocol is depicted in Figure 2.

To implement this protocol, we show two different possible implementations. In the first implementation, we chose the following HMI components:

- Tell-tale light showing the ADS view of preferred mode
- Push-button to for the manual driver to ask for mode change (first action)
- Tell-tale light showing whether the ADS is prepared for a change as requested by the manual driver
- Lever for the manual driver to select mode (second action)

Any failure mode of these four HMI components then needs to be included in the safety analysis, and this in combination by any single mistake by the manual driver.

To summarize, a fault-free uninterrupted transition from the MD mode to the AD mode in this example follow the steps:

- The manual driver drives the vehicle (MD mode)
- The ADS declares it is ready to take over by changing the preference tell-tale to AD mode available
- The manual driver asks to take over by pressing the push-button
- The ADS acknowledges that it is prepared by indicating the readiness tell-tale and unlocking the lever
- The manual driver changes the lever to AD mode position
- The ADS locks the lever, and continues to drive in AD mode

The transition from AD mode to MD mode is performed in a similar way, i.e., the manual driver may either independently, or suggested by the ADS, start by asking for a mode change. The ADS then acknowledges by indicating on the readiness tell-tale and unlocking the lever. Finally, the manual driver changes the lever to the MD mode position and starts to drive manually.

### *B. Safety Analysis*

In the following section, the above protocol and implementation is analysed with respect to its sensitivity to any human mistake, vehicle component failure, or a combination of these. Hence, we walk through the detailed state diagram and investigate the possible failure consequences at any state. When doing the safety analysis, we document the result in Table I. The columns are:

- Protocol state
- HMI failure to investigate
- Possible driver mistake
- Consequence in words
- Consequence in terms of safe/unsafe

Each row in this table marked as unsafe in the last column, needs to be protected by a corresponding safety requirement allocated to restrict this HMI failure. If all occurrences of an unsafe consequence are protected by appropriate safety requirements, the protocol implementation is deemed safe. For the safety argumentation to be valid, it is important that the table is shown to be complete. This includes an argumentation that all possible human mistakes are considered.

### C. Safety Assessments

As concluded from the safety analysis in Table I, there are four ways for this first example protocol implementation to fail in an unsafe way, caused by either of a manual mistake, a vehicle component failure, or a combination of these. The four failures that we need to avoid maintaining safety are:

- The ADS cannot correctly sense the mode lever position, which may cause *mode confusion*.
- The ADS cannot guarantee lock of the mode lever according to the protocol. This in combination with the manual driver moving the mode lever to AD mode, without noticing it, may cause *mode confusion* (or *unfair transition* if discovered by the manual driver).
- The ADS cannot guarantee locking of the mode lever according to the protocol. This in combination with the manual driver changing lever position from MD to AD, without getting acknowledgment of a prepared ADS, may cause *unfair transition*.
- The ADS cannot guarantee unlocking the mode lever according to the protocol. Which may cause *stuck in transition*.

TABLE I. SAFETY ANALYSIS OF TRANSITION PROTOCOL

| Protocol state | HMI failure | Driver mistake | Consequence | Safe/ Unsafe |
|---|---|---|---|---|
| MD - normal drive | Fault in lever lock | No | MD driver not trying to touch lever. Stay in MD. | Safe |
| MD - normal drive | Fault in lever lock | Driver changes lever position without asking for change first. | Unfair transition. | Unsafe |
| MD - normal drive | Fault in preference tell-tale | Any mistake or correct behaviour | MD cannot change locked lever. Stay in MD- normal drive. | Safe |
| MD - AD available | Fault in lever lock | No | MD driver not trying to touch lever. Stay in MD. | Safe |
| MD - AD available | Fault in lever lock | Driver changes lever position without asking for change first. | Unfair transition. | Unsafe |
| MD - AD available | Fault in preference tell-tale | No | Stay in MD | Safe |
| MD - AD available | Fault preference tell-tale | Driver ignores lack of availability | Transition sequence fulfilled. Change to AD. | Safe |
| MD - requested AD | Fault in push-button | Any mistake or correct behaviour | No Acknowledge by AD. Lever still locked. Stay in MD. | Safe |
| MD - prepared AD | Fault in prepared tell-tale | Driver correct: Driver stops transition sequence | Time-out in protocol. Stay in MD. | Safe |
| MD - prepared AD | Fault in prepared tell-tale | Driver incorrect: Driver ignores lack of ack. | Transition sequence fulfilled. Change to AD | Safe |
| MD - prepared AD | Fault in lever lock | Driver correct: Driver tries but cannot fulfil transition sequence. | Stuck in transition. | Unsafe |
| MD - prepared AD | Fault in lever lock | Driver incorrect: Driver doesn't continue transition sequence. | Time-out in protocol. Stay in MD. | Safe |
| AD – taking control | Fault in lever sensor | Any mistake or correct behaviour | Mode confusion | Unsafe |
| AD – normal drive | Fault in lever lock | No | MD driver not trying to touch lever. Stay in MD. | Safe |
| AD – normal drive | Fault in lever lock | Driver changes lever position to MD without asking for change first, and without noticing what is happening. | Mode confusion. (Unfair transition, if realized later). | Unsafe |
| AD – normal drive | Fault in preference tell-tale | No | MD acts as in normal AD mode. Stay in AD or ask for transition. | Safe |
| AD – normal drive | Fault in preference tell-tale | Driver tries to change lever position but it is locked in AD position. | Stay in AD. | Safe |
| AD – asking for MD | Fault in lever lock | No | MD not touching lever without asking for change first. Stay in AD. | Safe |
| AD – asking for MD | Fault in lever lock | Driver changes lever position by mistake without noticing it in the first place, and without asking for change first. | Mode confusion (Unfair transition, if realized later). | Unsafe |
| AD – asking for MD | Fault in preference tell-tale | Any mistake or correct behaviour | MD can request MD mode or stay in AD mode. | Safe |
| AD – requested MD | Fault in push-button | Any mistake or correct behaviour | No Acknowledge by AD. Lever still locked. Stay in AD. | Safe |
| AD – prepared MD | Fault in prepared tell-tale | No | Driver stops transition sequence. Time-out in protocol. Stay in AD. | Safe |
| AD – prepared MD | Fault in prepared tell-tale | Driver ignores lack of ack. | Transition sequence fulfilled. Change to MD | Safe |
| MD – taking control | Fault in lever lock | Any mistake or correct behaviour | Driver tries but cannot fulfil transition sequence. Stuck in transition | Unsafe |
| MD – taking control | Fault in lever sensor | Any mistake or correct behaviour | Mode confusion | Unsafe |

As we assume that the manual driver may make any single failure at any time, the way to argue for avoiding the above failures is to put the entire responsibility on the ADS. This implies that we put three safety requirements on the HMI of the ADS:

- ASIL D on restricting faulty lever sensor, i.e., the lever sensor needs to be always correct.
- ASIL D on restricting lever lock faulty unlocked.
- ASIL D on restricting lever lock faulty locked.

If we can guarantee that the ADS HMI is implemented according to these three safety requirements, we can claim that we make a safe transition even in the presence of an arbitrary single manual mistake. This takes care of all three aspects (*mode confusion*, *unfair transition* and *stuck in transition*) of a safe transition.

If ASIL D sensors and/or ASIL D locks are considered either unavailable or very expensive, we may consider redundancy implementation techniques. Instead of one sensor always telling the correct lever position with ASIL D attribute, we may consider three (sic!) sensors each with ASIL B. If at least two of the three are correct, we can stay safe. This means that we need to restrict that two of the three are failing. This shall be guaranteed with a total ASIL D, which we distribute as ASIL B on each sensor. Similarly, using ASIL A sensors would require seven times redundancy. If four out of seven are working we consider it as safe. This means that we need to restrict that four of the sensors are failing. This shall be guaranteed with a total ASIL D, which we distribute as ASIL A on each sensor.

As a second example in this paper, we use the same protocol, but chose other means of HMI components. Instead of pushing a button to initiate the AD->MD transition, the driver keeps his eyes focused on the traffic on the road for some seconds. Instead of a tell-tale on the instrument cluster to indicate to the driver that a mode change is prepared, we use a heads-up display (HUD) icon. Finally, instead of a lever for the driver to indicate the mode change, we choose a flip of some lateral steering wheel segments. These means can be argued as in-line with the idea of not distracting the driver, but making sure that while performing the transition sequence, the driver keeps attention to the driving task as well.

When we perform a similar analysis of this protocol implementation as we did for the first example in Table I, we will get exactly the same results. This means that in presence of fault-free HMI components, the implemented protocol is robust to any single driver mistakes. Furthermore, to guarantee safety in the presence of HMI component failures and any single manual mistake, we need to put safety requirements restricting failures on the steering wheel flip indicator, and on the locking mechanism of the steering wheel flipping mechanism.

Both the above examples can be argued as idiotic implementations. The idea here is not to show what is the best implementation of a protocol, but rather to illustrate the technique to investigate the failure modes of the HMI components together with the possible manual mistakes according to a certain protocol.

In the last example, we go one step further and construct a protocol tolerant to any double human mistakes. This implies that any handover sequence involves three coordinated actions by the driver (two actions could be a double failure).

The example protocol can be found by just expanding the previous one with one action from each part, as shown in Figure 3.

The chosen HMI components for the sake of the argument are the following:

- A tell-tale + sound indicating to driver that the ADS prefers to leave control to the Driver.
- A button for the Driver to push when initiating a mode change from AD to MD.
- A HUD icon asking the driver to show readiness to take over control.
- An eye tracker checking that the Driver keeps the eyes on essential parts of the road and traffic environment.
- A HUD icon telling the Driver that it is OK to take over control.
- A steering wheel with flippable lateral segments for the driver to indicate mode of driving (when compressed in AD mode; when expanded in MD mode).
- A locking mechanism, making sure the steering wheel segments only are flipped at valid moments according to the protocol.
- A HUD icon telling Driver when mode change from MD to AD is available.
- A button for the driver to press when initiating a MD to AD hand over.
- A HUD icon asking for confirmation from Driver that mode change to AD is intended.

We can now again walk through the protocol and investigate the failure modes in similar was as was done in Table I. This leads to similar result following the general pattern:

*Every ADS HMI component being responsible for either of*

- *not allowing the driver to (by mistake) perform such a forward transition in the protocol that makes any of the user actions unneeded to complete the transition*
- *not misinterpret the driver actions such that a forward transition in the protocol that makes any of the user actions unneeded to complete the transition*
- *not hindering the driver to perform the last action in the transition protocol, when the user should expect it to be aloud*

*will get an ASIL requirement. And only those.*

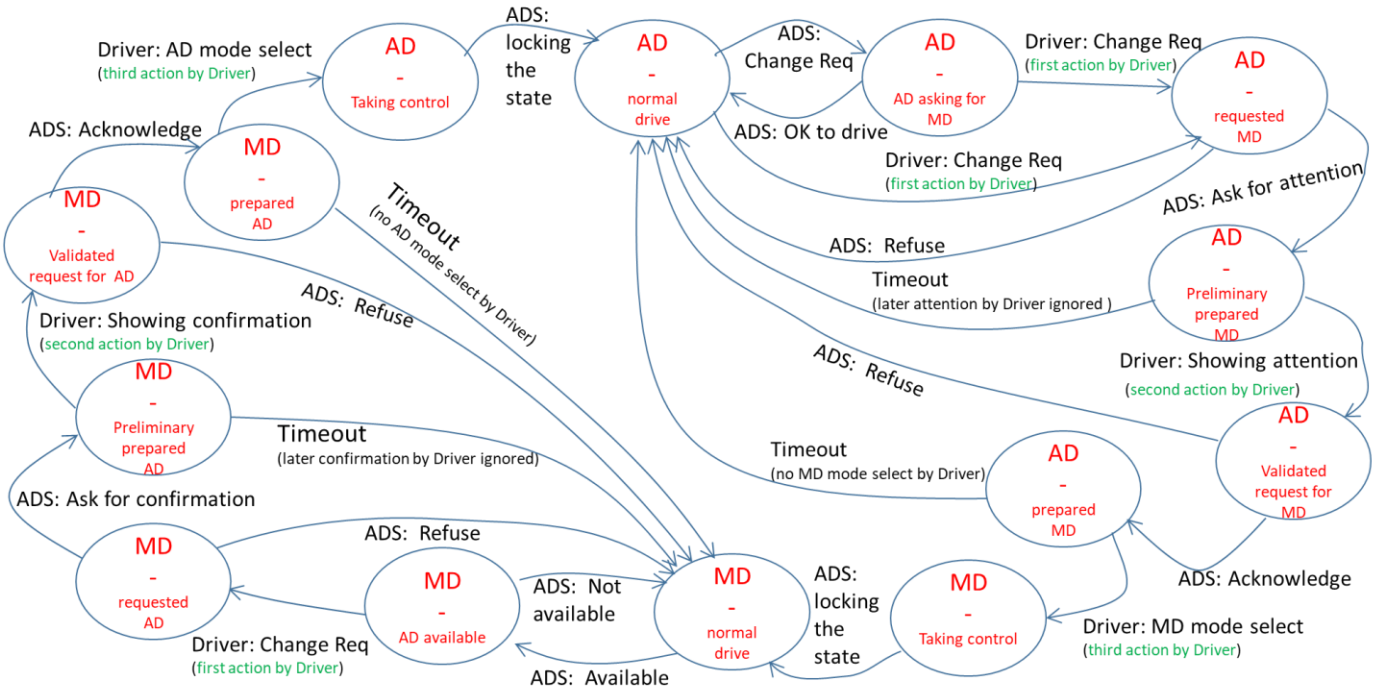


Figure 3. Example of an elaborated 3-action protocol, robust against two manual failures.

The first two conditions above are to guarantee absence of *mode confusion* and of *unfair transitions*, and the third one is for avoiding *stuck in transition*. Note that what is listed above are just the safety requirements on the HMI components. The internal logic executing the protocol is still subject to safety requirements for any failing transition.

## VI. CONCLUSION

When introducing an automated driving system (ADS) according to SAE Level 4, which takes full responsibility to drive the vehicle once activated, it becomes crucial to ensure safe transitions between the manual driver and the ADS. The existence of dual driving modes brings three new sources of risk, namely *unfair transition*, *mode confusion* and *stuck in transition*.

We propose to define a safe transition as a transition where neither a (complex) manual mistake nor an E/E failure, nor combination of these, leads to an *unfair transition*, *mode confusion* or *stuck in transition*.

We do not prescribe what manual failures to consider, but rather showing a methodology how to perform safety analysis of implementations of transfer protocols. Given that we agree on what set of single, double or multiple failures by the driver to assume, we show how to argue that an appropriate set of safety requirements on the HMI components would be sufficient to deem the HMI of the ADS functionally safe.

Furthermore, we demonstrate on some system examples how to allocate safety requirements on HMI elements to ensure safe transitions, and we show how the same protocol can be implemented by different HMI components. We also show an example of a protocol and implementation designed to be robust to dual driver mistakes.

Results from this example show that it is sufficient to allocate safety requirements on the sensor of the driver action and of the lock of the mode control, to ensure a safe transition. No safety requirements are needed on visual feedback to the driver, e.g., displays. We remark that the example implementations by no means are unique or optimal solutions to the safe transitions problem, but intended to illustrate a methodology.

## ACKNOWLEDGMENT

This research has been supported by the Swedish government agency for innovation systems (VINNOVA) in the ESPLANADE project (ref 2016-04268).

## REFERENCES

[1] R. Johansson, J. Bergenhem, and M. Kaalhus, "Safe transitions of Responsibility in Highly Automated Driving," DEPEND 2016: The Ninth International Conference on Dependability, July 2016.

[2] SAE, "Surface Vehicle Recommended Practice – J3016 – Taxonomy and Definitions for Terms Related to Driving Automation systems for On-Road Motor Vehicles", September 2016.

[3] C. Gold, D. Damböck, K. Bengler, and L. Lorenz, "Partially Automated Driving as a Fallback Level of High Automation," 6. Tagung Fahrerassistenzsysteme. Der Wig zum Autom. Fahren., 2013.

[4] National Highway Traffic Safety Administration, "Human Factors Evaluation of Level 2 And Level 3 Automated Driving Concepts Past Research, State of Automation Technology, and Emerging System Concepts," http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2014/812043_HF-EvaluationLevel2andLevel3AutomatedDrivingConceptsV2.pdf, retrieved: June 2016.

[5] M. H. Martens and A. P. Van Den Beukel, "The road to automated driving: Dual mode and human factors considerations," IEEE Conf. Intell. Transp. Syst. Proceedings (ITSC) , 2013, pp. 2262–2267.

[6] F. Naujoks, C. Mai, and A. Neukum, "The effect of urgency of take-over requests during highly automated driving under distraction conditions," Adv. Hum. Asp. Transp. Part I, vol. 7, July 2014, p. 431.

[7] M. Blanco, J. Atwood, H. M. Vasquez, T.E. Trimble, V.L. Fitchett, J. Radlbeck, and J. F. Morgan, "Human factors evaluation of level 2 and level 3 automated driving concepts," Report No. DOT HS 812 182, Washington, DC, National Highway Traffic Safety Administration, 2015.

[8] N. Merat, A.H. Jamson, F. F. C. H. Lai, M. Daly, and O. M. Carsten, "Transition to manual: Driver behaviour when resuming control from a highly automated vehicle," Transportation Research Part F: Traffic Psychology and Behaviour, 26, 1–9, 2014.

[9] T. Helldin, G, Falkman, M. Riveiro, and S. Davidsson, "Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving," In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13 (pp. 210–217). New York, New York, USA: ACM Press, 2013.

[10] J. Beller, M. Heesen, and M. Vollrath, "Improving the Driver-Automation Interaction: An Approach Using Automation Uncertainty," Human Factors: The Journal of the Human Factors and Ergonomics Society, 55(6), 1130–1141, 2013.

[11] B. Seppelt, and J. Lee, "Making adaptive cruise control (ACC) limits visible," International Journal of Human-Computer Studies, 2007.

[12] S. Brandenburg and E. Skottke, "Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving?," IEEE 17th Int. Conf. Intell. Transp. Syst. (ITSC), pp. 2978–2983, 2014.

[13] A. Marinik, R. Bishop, V. Fitchett, J. F. Morgan, T. E. Trimble, and M. Blanco. "Human factors evaluation of level 2 and level 3 automated driving concepts: Concepts of operation," Report No. DOT HS 812 044. Washington, DC: National Highway Traffic Safety Administration., July 2014.

[14] H. Orlady, and R. Barnes, "A Methodology for Evaluating the Operational Suitability of Air Transport Flight Deck System Enhancements," SAE Technical Paper # 975642, 1997.

[15] T. Inagaki, "Design of human–machine interactions in light of domain-dependence of human-centered automation," Cognition, Technology & Work, Volume 8, Issue 3, pp 161-167, 2006

[16] T. Lackman, "Utredning och kartläggning av tillfällen då människan räddat och förbättrat en situation där automatiken inte räckt till eller fungerat fel," Strålskerhetsmyndigheten 2011:24, ISSN 2000-0456, 2011.

[17] J. Cunningham "Break the monotony," Professional Engineering, 20(20), 33-33, 2007.

[18] D.B. Kaber, and L. J. Prinzel, "Adaptive and adaptable automation design: A critical review of the literature and recommendations for future research," NASA/TM-2006-214504, September 2006.

[19] R. Johansson, C. Bergenhem, and H. Sivencrona, "Challenges of Functional Safety in ADAS and Autonomous Functions," SAE World Congress, Detroit, April 2014.

[20] J. A. Michon, "A Critical view of driver behavior models: What do we know, what should we do?,",in L. Evans & R.C Schwing (eds.) Human behavior and traffic safety (pp. 485-520). New York: Plenum Press, 1985.

[21] R. Sukthankar, "Situation Awareness for Tactical Driving," Ph.D. thesis, Robotics Institute, Carnegie Mellon University, USA, January 1997.

[22] T. X. P. Diem and M. Pasquier, "From Operational to Tactical Driving: A Hybrid Learning Approach for Autonomous Vehicles," 10th Intl. Conf. on control, Automation, Robotics and Vision, Hanoi, Vietnam, December 2008.

[23] ISO, "International Standard 26262 Road vehicles -- Functional safety", November 2011.