Transportation Research Part F 43 (2016) 279–301

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Supporting the changing driver's task: Exploration of interface designs for supervision and intervention in automated driving

Arie P. van den Beukel [*,1], Mascha C. van der Voort [2], Arthur O. Eger [3]

Human Centred Design Group, PO Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 2 February 2016
Received in revised form 10 June 2016
Accepted 3 September 2016
Available online 28 September 2016

Keywords:
Automated driving
Partial automation
Driver-vehicle interface
Supervision
Intervention

ABSTRACT

Driving automation leads to a changing role for drivers, that is from manual vehicle control to supervising automation. Supervision of partial automation requires now and then intervention. Since the automation causes low vigilance and out-of-the-loop performance problems, this changing role is not well suited for human operators. To explore how driver-vehicle interfaces can support drivers in their changed role, we tested three concepts. Concept A was a baseline reference, providing only acoustic warnings. Concept B presented status-information and warnings behind the steering wheel. Concept C used illumination and haptic feedback in the seat-pan to direct attention outside the vehicle and to stimulate response. Concept C only provided vibrotactile feedback when intervention was needed. Results of our study show improved support for supervision with the illumination-concept, i.e. better hazard-detection and raised levels of Situation Awareness in some scenarios relevant for supervisory control. Knowing that supervision will be the dominating driver's responsibility during partially automated driving, the illumination-concept is a recommended solution for support of the driver's changing role. Nonetheless, neither concept B, nor C, showed additional support for intervention compared to the baseline. It was hypothesised that the combination of concept C's stimuli for intervention-support caused counter-productive levels of annoyance. Furthermore, we concluded that intervention and supervision benefit from different interface-features and discussed possible causes underlying ambiguity between support for supervision and support for intervention shown with concept C. Therewith, the considerations in this paper contribute to further development of – and knowledge about – appropriate driver-vehicle interaction while vehicle-operation advances into operating partially automated driving systems.

1. Introduction

Automotive industry has started introducing vehicles to the market that allow automated driving (Ackerman, 2015; Daimler (Producer), 2015). Automated cars are assumed to raise comfort, cause less accidents, reduce congestions and decrease environmental impact of driving (EU, 2011; Infra & Milieu). In line with roadmaps for the development of automated driving, the introduced systems are defined as *partial automation*, meaning that the driver should permanently mon-

---

[*] Corresponding author.
E-mail address: a.p.vandenbeukel@utwente.nl (A.P. van den Beukel).

[1] Assistant Professor at University of Twente.

[2] Professor at University of Twente.

[3] Professor emeritus.

http://dx.doi.org/10.1016/j.trf.2016.09.009
1369-8478/© 2016 Elsevier Ltd. All rights reserved.

itor the automation and preserve final responsibility for safe driving (Bengler et al., 2014; Gasser & Westhoff, 2012; SAE_International., 2013). The reason behind this introductory level of automation and subsequent driver's responsibilities, is the complexity for an automated system to interpret or understand many contextual factors relevant to safe driving. Despite enormous technological achievements when introducing these systems, developers need to account for technological boundary conditions (Flemisch, Bengler, Bubb, Winner, & Bruder, 2014; Nieto, Otaegui, Vélez, Ortega, & Cortés, 2015). In fact, Nieto et al. (2015) indicated reliability of hardware and software as well as resistance to harsh conditions some of the key challenges for development of ADAS' sensors. The definition of partial automation assumes drivers are available at all time to take over control of the vehicle if required (De Winter, Happee, Martens, & Stanton, 2014; Gold, Damböck, Lorenz, & Bengler, 2013). In other words, the introduction of partial automated driving assumes that this capability is used in a way so that the driver is always sufficiently in the loop to take over control if requested by the automation or, more challenging, if necessary due to a hazard not perceived by the automation (Flemisch et al., 2014; Merat, Jamson, Lai, Daly, & Carsten, 2014; Merat & Lee, 2012; Toffetti et al., 2009). Obviously, the core difficulty with driver's interaction with partial automation is that it *assumes* driver's availability, though it is more likely that drivers will shift their attention to non-driving related tasks, while being placed out-of-the-loop. Consequently, during automation the driver's task changes from actively operating the vehicle to supervising the system (Martens & Van den Beukel, 2013; Merat, Jamson, Lai, & Carsten, 2012). However, supervision is not something humans are particular good at (Brookhuis, De Waard, & Janssen, 2001), due to low vigilance and behavioural adaptation, causing for example slower reaction times (Larsson, Kircher, & Hultgren, 2014), reduced attention to tasks (Farrell & Lewandowsky, 2000), or misinterpretation (Martens, Pauwelussen, Flemisch, Merat, & Caci., 2008; Merat et al., 2014). Nonetheless, supervision will, during automation, be the dominating driver's task and acceptance of partially automated driving is expected to depend heavily on the appraisal whether supervision (including now and then retrieving control) is less effort taking than a human driving himself or herself (Payre, Cestac, & Delhomme, 2014; Van den Beukel, Van der Voort, & Eger, 2015).

To summarise, despite the potential advantages of automated driving, partial automation brings the driver into a troublesome situation: being fully responsible for safe driving while having reduced authority and control (Bainbridge, 1983; Flemisch et al., 2014). Carefully designed driver-interfaces are therefore needed to support drivers in their changed role to supervise the automation and to support them retrieving control safely and adequately when required (Hoc, Young, & Blosseville, 2009). Existing research however, often focuses on the relation which individual aspects have with a specific task, to a lesser extent on a holistic view of how driver performance is influenced by a combination of interface-aspects applicable in practice (Nirschl, 2007; van Waterschoot & van der Voort, 2009). For instance, adequate support for the changing driver's role asks for improving the cognitive understanding of how a system reacts to different situations (i.e. support for supervision) and at the same time asks for improving the operational capabilities for reacting fast and adequately when solving a critical situation (i.e. support for intervention) (Van den Beukel & Van der Voort, 2016b). Interface-aspects that show support for these subtasks individually will not necessarily withstand when combined in one overall interface. This research therefore aims at: (a) exploring potential solutions of driver-interfaces that integrate both support for supervision and intervention, and (b) evaluating these interfaces within circumstances that are representable for the complex practice in which partially automated driving will be applied. To this pursuit, the research is based on a combination of exploration with an experiment-based evaluation of proposed concepts. The article starts with reviewing interface-aspects that are important for the development of driver-support. Next, two interfaces are being composed and compared to a baseline concept in a driving simulator test. The simulator's testing conditions and applied measurement aspects are adopted from a previously defined framework (Van den Beukel & Van der Voort, 2016b; Van den Beukel et al., 2015). Therewith, this research allows to assess the proposed concepts with regard to support for supervision and intervention. Based on comparison of these results with existing studies, the paper discusses to what extent the applied interface-aspects are beneficial to create desirable support. This paper concludes with recommendations for the application and further development of interfaces, which are supportive to the driver's changed role when driving automatically.

1.1. Behavioural improvements of raised LOAs

Despite the raised concerns with respect to the driver's changed role when utilising partially automated vehicles, some positive behavioural change or compensating adaptation is being associated with application of raised levels of automation (LOAs):

- Drivers who have adapted to the benefits of raised comfort (provided by the automation) also demonstrate more homogenous traffic participation, resulting in safer and more efficient driving, e.g.; less lane changing and reduced over-taking (Jamson, Merat, Carsten, & Lai, 2013; Van Arem, Van Driel, & Visser, 2006; Van Driel, 2007).
- If drivers are motivated or instructed to take benefit from the freed resources (due to partial automation) and use it for observing objects in the environment, higher levels of automation can result in improved situation awareness compared to manual driving (De Winter et al., 2014).

- Drivers who are familiarised with the technical boundary conditions of their assistance system, have higher awareness of how increased traffic density, or aggravated road-traffic circumstances, influences system performance. These drivers show the capability to compensate with their attention-level (e.g. reduce their involvement in non-driving related tasks) for impaired system-performance (Donges, 1999; Gartenberg, McCurry, & Trafton, 2011; Jamson et al., 2013; Merat et al., 2012).
- If expectations are aligned (e.g. drivers correctly anticipating system-state and -changes in system-state), then considerable improvement in drivers' ability to regain control is demonstrated (Larsson et al., 2014; Merat et al., 2014). Likewise, an expected stimulus during driving causes faster reaction times and is associated with more correct responses than an unexpected stimulus (Martens, 2004).

Besides, with prior warning, collision avoidance during automated driving situations (i.e. driver out-of-the-loop) is as good as during manually driving (driver in-the-loop) (Gold et al., 2013). This emphasises the importance to support drivers with their changed role (Merat et al., 2014) and means that drivers should be enabled to take their responsibility and intervene when required. Flemisch et al. (2014) formulated the required interaction in terms of: "drivers should have appropriate insight, authority and control".

1.2. Supervision with now-and-then intervention

This title characterises the changing driver's role for which we want to develop support in terms of improved driver-vehicle interaction. Proposed solutions often address supervision or intervention separately and usually depart from theoretical considerations. Nonetheless, we want to explore and test potential solutions that combine support for supervision and intervention in single concepts, which are applicable for implementation within approximately five years. Therefore, we now review important considerations and proposed solutions from previous studies. In general, existing approaches differentiate between either leaving drivers involved in the control-loop or to have them back in the control-loop as fast and effectively as possible. See for an extensive explanation of the concerns and recommendations related to out-of-the-loop performance: Bainbridge (1983), Endsley and Kiris (1995), Brookhuis et al. (2001), and Martens and Van den Beukel (2013).

1.2.1. Continuous information on system-state

Automation-induced accidents are often caused by lack of automation-feedback (Norman, 1990). Besides lack of operator's knowledge about the operational envelope of a system, also difficulties to distinguish between the technical characteristics of similar but not identical systems, are often caused for operator's misunderstanding and subsequent failure to respond correctly (Martens, 2007). An extensive review of 32 studies measuring effects of automated driving on situation awareness (De Winter et al., 2014), confirms that a proper feedback system could alleviate much of the concerns in operability of partially automated driving. Therefore, researchers advise to provide continuous feedback on system-state, like providing information on activation, deactivation, availability and malfunction (Martens & Van den Beukel, 2013; Nielsen, 2005). Then, proposals are often based on visual interfaces providing feedback on system's conditions and changes in system-state through a display. Although an on-road study by Stanton, Dunoyer, and Leatherland (2011) and a driving simulator study by Seppelt and Lee (2007) provide evidence that a visual display with continuous information on system-state improves driver's situation awareness, some studies have also raised concerns that the provided information could become a distraction. A continuous display of detailed capabilities and operational envelopes could result in display clutter and consequently confusion and distraction (Martens et al., 2008). The thoughts behind providing continuous information on system-state are to enable drivers to return more easily in the control-loop, because it provides information on the parameters that are important for perception, decision-making and response - the basic elements within a control-loop.

1.2.2. Timely warnings

Other studies explore support in terms of providing warnings that allow drivers to turn back in the control-loop on time and to demonstrate effective intervention. Gold et al. (2013) showed that providing a take-over request (i.e., a simple audio-visual warning) 5–7 s in advance ensures that drivers of a highly automated car avoided a stationary object, even if they were not attending to the road prior to the take-over request. Van den Beukel and Van der Voort (2013) found that drivers who were out-of-the-loop and travelling with time headways of about 1.5 s (which are during manually driving considered 'safe'), failed in almost half of these occasions (47.5%) to avoid collision with a broken down vehicle during an unexpected event. However, this study also showed that providing an improved level of situation awareness (through e.g. appropriate system feedback and timely warning) improves the ability for successfully taking back control.

1.2.3. Cooperative control

Many studies have addressed possible support of the intervention-task by keeping operators somehow involved in the control-loop (Abbink & Mulder, 2010; Flemisch et al., 2014; Hoc et al., 2009). Referring to sequences of task performance involving perception, decision-making and implementation, studies demonstrated that operator's ability to intervene and retrieve control is considerably better when being involved in the *implementation*-step (Endsley & Kaber, 1999; Endsley & Kiris, 1995; Kaber & Endsley, 1997). In contrast, if the automated system failed, the removal of human control from the operational part resulted in inability to recover (Kaber & Endsley, 2004). Research of how levels of automation influence operator's controllability often uses generic devices from the computer-domain. However, Flemisch et al. (2014) proposed an advanced stick-controller – based on haptic input and feedback, specifically for the automotive domain. The device enables drivers to remain involved in the operating-task and at the same time delegates automation depending on the circumstances. Also Mulder, Van Paassen, and Abbink (2008) as well as Mulder, Abbink, and Boer (2012) explored haptically enhanced automotive-devices, like haptic force-feedback in the steering wheel and a haptic pedal. They concluded that haptic control enabled the driver to remain in the loop with enhanced performance at reduced control activity. Therewith the haptic control enabled to take advantages of raised comfort through automation while mitigating the controllability issues of full automation.

1.2.4. Acceptance & comfort

We acknowledge that besides effectiveness of the desired support, the acceptance of provided support is also important. In addition to possible distraction due to the transferred information, other negative side-effects need to be avoided too. Among these, it certainly needs to be accounted for that stimuli might become annoying or cause startle-reactions (Baldwin & May, 2011; Blumenthal, 1996; Fagerlonn, 2011).

1.2.5. Interface's scope

Based on previous findings we assume that a driver-vehicle interface for partially automated driving provides appropriate support for supervision and intervention if it is characterised as follows:

> *The interface directs attention to the location where attention is needed with "correct" level of urgency. In addition, the interface provides appropriate feedback and feedforward with respect to foreseen and likely changes in system state without causing counter-productive effects, like distraction. Moreover, the interface enhances fast and effective human intervention if required.*

If possible, we want to include advantages of previous findings within concepts that combine support for supervision and intervention. The next section therefore reviews how different modalities influence afore considerations. The goal is to compile different concepts that demonstrate potential advantages through comparison within meaningful circumstances and are realistic with respect to implementation within approximately five years.

1.3. Considering modalities for concept development

When considering appropriate modalities for development of the interface, we need to realise that the information we want to transfer does not only convey warning messages but also more content-rich (and often less urgent) information like status information. Furthermore, to weigh pros and cons of different modalities we need to understand that signal effectiveness is related to perceived signal's *urgency, annoyance* and *semantics* (Baldwin & Lewis, 2014; Baldwin & May, 2011). In other words, modality alone does not drive differences in signal effectiveness.

1.3.1. Audible interfaces

In general, auditory signals are recommended as a base attention retrieving signal especially for urgent situations, like primary collision avoidance warning (Kiefer et al., 1999; Tan & Lerner, 1995). Auditory alarms are especially supportive to retrieve attention from distracted drivers (Lee, Suh, & Benbasat, 2001). Concerns for an auditory interface are that they are more prone to annoyance and auditory signals receive often competition from In-Vehicle Information Systems (IVIS), like radio and navigation.

1.3.2. Visual interfaces

Visual interfaces have the advantage to enable more content-rich transfer of information. Besides, visual interfaces enable users their own pace of information retrieval. Often, perception of the signal can be interrupted and continued later on. In this respect (self-paced, pausing) a visual interface combines well with the visually dominated and highly changeable driving task. Concerns for a visual interface are as follows: (a) The information can become a distraction (as explained in the previous section), and (b) Visual signals can be easily overlooked within the stimulus-rich visual driving environment (Martens, 2007).

1.3.3. Tactile interfaces

It is intrinsic to tactile interfaces that they hold, in comparison to visual interfaces, a high base level of perceived urgency. This makes tactile interfaces less suitable to convey supervision-information. Compared to visual feedback, this modality also has a reduced resolution to convey a multitude of potentially relevant messages (Fitch, Hankey, Kleiner, & Dingus, 2011). Haptic alerts are especially effective for cueing distracted drivers' attention back to the road (Green et al., 2008; Ho, Tang, & Spence, 2005; Lee, Hoffman, & Hayes, 2004). As an attention signal, tactile feedback has in this respect a comparable function as audible signals. In addition, a tactile stimulus can also help to select and speed-up correct control (Fitch et al., 2011; Flemisch et al., 2014). Due to the strong link of tactile stimulus with the neuro-muscular system tactile interfaces could help with affordance for action (Smisek, Van Paassen, Mulder, & Abbink, 2013). Therefore tactile interfaces allow creation of combined devices for both feedback and control (Abbink & Mulder, 2010).

1.3.4. Signal urgency

With respect to perceived signal's urgency, there is a fundamental relationship between the signal's intensity (frequency, wavelength, pace and duration, etc.) and perceived urgency. For example: If the frequency of a sound increases, and/or as the time interval between pulses of sound decreases, it is perceived as increasingly urgent (Edworthy, Loxley, & Dennis, 1991; Hellier & Edworthy, 1999). Likewise, as the wavelength of visible light increases (hue or perceived colour changing from green to yellow to red) it is perceived as more urgent (Chapanis, 1994; Wogalter, Conzola, & Smith-Jackson, 2002). The behaviour of tactile stimuli is in this respect comparable to sound (Baldwin & Lewis, 2014), meaning that if the frequency and/or pulse of a vibration increases, this is perceived more urgent. Furthermore, there is a direct relationship between perceived urgency and annoyance, such that as a signal becomes more urgent it is also perceived as more annoying (Baldwin & May, 2011; Marshall, Lee, & Austria, 2007). However, the context in which the signal is presented influences this relationship (Wiese & Lee, 2004). More urgent signals are perceived as less annoying in conjunction with situations where the high urgency seems appropriate (collision warnings) relative to situations where it is less appropriate to receive a very urgent signal (e.g., navigation command or email alert) (Marshall et al., 2007).

1.3.5. Modality influences relation between urgency and annoyance

Baldwin and Lewis (2014) has researched the trade-off between annoyance and perceived urgency for signals with different modality. Their research showed that especially visual colour and tactile intermitting modality are potentially useful, because the characteristics of these modalities can be changed to provide more urgency, with a minimum effect on increasing annoyance (indicated by the steep slope of the urgency-annoyance lines of these signals in Fig. 1). The research also showed that the visual colour interface has lowest base annoyance level and that the tactile interface has a higher baseline urgency level than a visual colour interface. The range-lengths and base-levels also provide insight on the stimuli applicability to provide different signals. For example, a stimulus with long range (e.g. tactile intermitting) could be manipulated to create a signal with moderate urgency (like an attention-alert) as well as high urgency (e.g. intervention-signal). However, the applicability is limited by threshold-levels of urgency: When urgency-levels are too high, their effectiveness for intervention-support (like crash avoidance) reduces due to startle effects on the driver (Baldwin & May, 2011; Blumenthal, 1996). For further utilisation of results from the Baldwin and Lewis (2014) study, it needs to be mentioned that the stimuli were single modal (i.e. different modalities were not tested simultaneously as a combined stimulus) and the stimulus was intended as attention-signal, not compiled to bring information-rich contents across.

1.3.6. Semantic factors

Available data on the semantics of stimuli relate to utilisation of words and colour. In line with expectations, the highest wavelengths visible to the human eye (i.e. 600–750 nm; orange and red) are associated with high urgency (Braun, Sansing, & Silver, 1994b). Written words may express different levels of urgency, like "Danger" compared to "Notice" (Wogalter, Kalsher, Frederick, & Magurno, 1998). Likewise, different semantics of spoken words allow differentiation in perceived hazard. Although we favour language-independent solutions, it is important for our interface-development to acknowledge that semantic factors (e.g. signal word) interact with circumstances (e.g. loudness) at which an auditory warning is made heard (Baldwin & May, 2011; Edworthy, Hellier, Walters, Clift-Mathews, & Crowther, 2003) and that this interaction can impact both perceived urgency and collision avoidance response. A typical example is provided by Baldwin and May (2011) showing with a collision avoidance task in a driving simulator test that the majority of people crashed when either receiving no warning or receiving an acoustic warning with the most urgent signal word "Danger" at an high intensity (85 dB). It was hypothesised that the extremely urgent acoustic alarm caused startle and was therefore not successful in avoiding crashes. In contrast to audible interfaces, combining written signal words with colours in visual interfaces allow creation of more moderate levels of perceived urgency (Braun, Sansing, Kennedy, & Silver, 1994a). That is; combinations of word and colour with similar semantics (e.g. using the word "Danger" on a red background) are not reported to cause inappropriate high levels of urgency.

1.3.7. Multimodal interfaces

In view of the pros and cons associated with the different modalities, it is understandable that researchers recommend to not rely on one modality but explore multimodal interfaces (PReVENT., 2009). Multimodal interfaces bring advantages when different modalities are utilised in a complementary fashion. For example, vision exists over time and sound exists over space (i.e. vision and sound are independent of respectively time and space). However, if too much information is provided information-overload occurs. Besides, users may have difficulties comprehending information coming from different modalities simultaneously, either disregarding or totally ignoring one or the other (Lee et al., 2001). Based on the considerations for the potential support provided by different modalities and the introduced vision on desired solutions that combine supervision and intervention, we have compiled two concepts with multimodal features and defined one base-concept for comparison. The concepts are explained in the next paragraph.

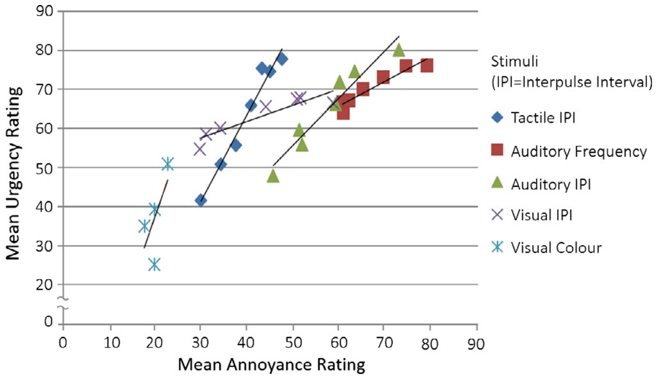


Fig. 1. Visualisation of relative impact of visual, auditory, and tactile modalities on perceptions of urgency and annoyance, adopted from Baldwin and Lewis (2014). The more vertical the line the greater the signal utility. That is: large increase in urgency can be achieved with minor increase in annoyance.

1.4. Concepts

The concepts' function was to provide signals to either attract driver's attention for potentially hazardous situations or to urge for intervention. The signals to ask for attention are referred to as 'soft warnings'. The signals intended to request take-over are called 'hard warnings'. The concepts' signals and stimuli were composed as follows:

- Concept A served as baseline, providing only audible signals. The audible soft warning was a high-frequency (approximately 1 kHz) "beep"-sound with a duration of 2.5 s, transcending engine-sound with about 4 dB. The hard warning consisted of an intermitting rattle-like sound at an intensity of 70 dB (roaring over engine noise with approximately 10 dB) and a duration of 2.0 s.
- The second concept (B) provided in addition to audible signals also feedback upon system-state with a graphical icon, displayed behind the steering wheel. The audible warnings were the same as with concept A. Examples of how the graphical icon depicted different soft and hard warnings are provided with the inserted images in Fig. 2.
- A third concept (C) offered the same audible warnings as A, but provided additional signals through illumination of the steering wheel and – occasionally – the windscreen. The illumination around the steering wheel represented the system to be active. Illumination at the sides of the windscreen or side windows was non-permanent. It lit op at the occasion of an event requiring either attention or intervention. The occurrence of illumination intended to attract attention away from non-driving tasks and the location of illumination intended to direct driver's focus towards the location outside the vehicle where attention was needed. Furthermore, the colour of the illumination conveyed urgency and helped to distinguish between soft and hard warnings. The blue illumination around the steering wheel was visible as long as system control was active and faultless. Yellow illumination (in the side-windows or windscreen) pointed at needed attention, for example merging-in traffic from an on-ramp. Hence, this configuration conveyed a "soft warning". Red illumination conceived danger and urged for direct intervention to take back control (i.e. a "hard warning"). In addition to visually directing attention, the hard warning of concept C provided tactile feedback. Vibrotactile pots were placed in the front-end of the seat cushion intended to create an affordance for appropriate intervention (e.g. braking). The tactile stimulus existed of two pulses of about 600 ms. The intended tactile affordance was similar to Fitch et al. (2011) who observed that a haptic seat helped distracted drivers to gain quick and accurate manual responses to a critical event. Similar findings were reported by Hoffman, Lee, and Hayes (2003).

Based on considerations derived from existing research – and explained in this section – the aim of our research was (a) to explore the possibilities for guiding focus towards relevant locus of attention outside the vehicle, (b) to adequately inform drivers about system-state, and (c) to create an affordance for action if intervention is needed. All, intended to support drivers in their changed role for supervision with now-and-then intervention. Clearly, the compilation of the concepts is not intended to test fundamental differences between interface modalities. In general, visual interfaces provide a large variety of different instantiations, like using colours, icons, screens with different possibilities for size and location. The visual characteristics between our concepts B (icon-interface) and C (illumination) are deliberately chosen to reflect different approaches: the icon-concept (B) allows to convey detailed information upon system-state and related changes in the road-environment (like detection of road-lines) – but assumingly requires more cognitive load for information-processing. The illumination-concept (C) is intended to guide attention outside the vehicle presuming to help drivers return in the control-loop. Although, other instantiations could provide similar effects, and based on practicability the number of concepts was limited to three. Moreover, a Heads-Up Display (HUD) was considered, because the projected image at focus-distance outside the vehicle seems promising to promote driver's focus at the road scenery (Götze & Bengler,

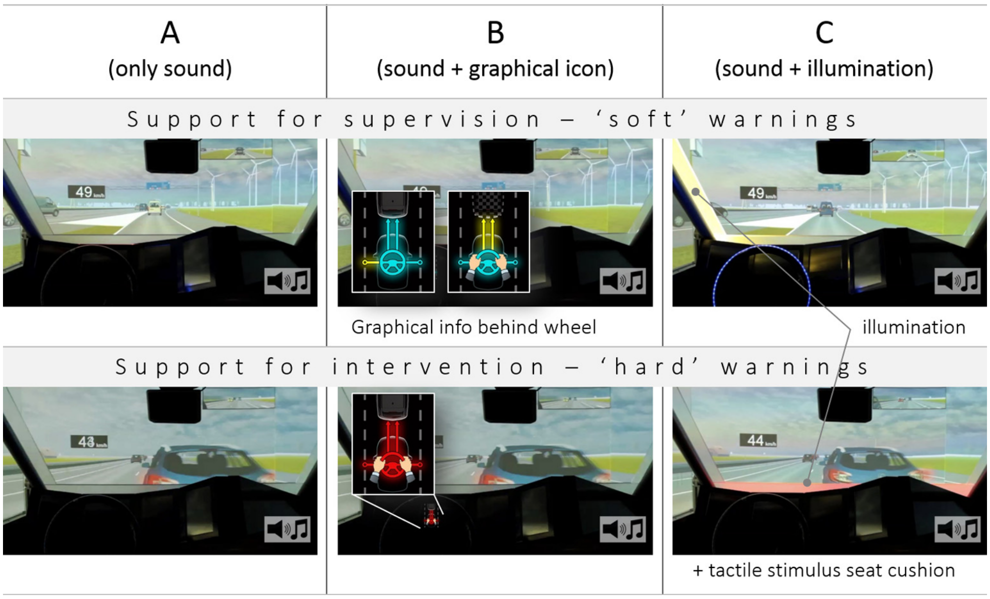


Fig. 2. Concepts providing different combinations to utilise modalities for providing support for supervision and intervention. Concept A intended a baseline concept providing only audible warnings.

2015). However, the limited viewing-angle of current systems and subsequent inability to connect traffic-information from a variety of road-locations via a HUD to the driver, caused this technology to be excluded from our current test.

2. Method

The previous chapter reviewed interface-aspects important to support drivers with their changed role (i.e. supervising the automation with now-and-then intervention) and introduced potential concepts. The aim of this research was to propose solutions that demonstrate desired support. Therefore the introduced concepts will now be tested within circumstances representative for partially automated driving. This chapter describes the applied within-subjects experiment design using a mid-fidelity driving simulator. While providing the results of this experiment, the next chapter assesses the proposed concepts with regard to support offered for supervision and intervention. Based on comparing results from this research with existing studies, chapter 4 will then discuss the influence of the reviewed interface-aspects on desired support. Thereafter, the last chapter concludes with recommendations on the identified interface-aspects that are beneficial to create the desired support for the driver's changing role when driving automatically.

2.1. Participants

Thirty-seven persons participated and their age ranged from 20 to 64 years (Mean = 47.0, SD = 12.6). Seventeen female participants took part and also twenty males. All participants had normal or corrected to normal sight. Participants were recruited using announcements in local newspapers. None of the participants had experience with using automated driving from earlier driving simulator experiments. This was to avoid prior knowledge of the traffic scenarios and circumstances to be expected during the experiment. Other than possession of a driving licence for a minimum of two years and maximum age of 65 years, no particular criteria were used for recruiting participants. Participants did not receive compensation, but among participants three nation-wide accepted shopping-vouchers were raffled off as an incentive. The characteristics of the participants were diverse, like students, professionals, unemployed and retired people. Though participants were interested in the subject of automated driving, their base attitude was diverse too: Some people's expectations towards the technology run high, and others were more in general interested in automobiles and car-driving. Another considerable part of the participants had extrinsic motivation to participate since they considered our research to be very relevant for society. The occupation among participants was diverse too, like salesman, teacher, businessman, pilot, engineer and journalist.

2.2. Scenarios & system description

During the experiment, the concepts were tested within a total of six different scenarios, see Table 1. In order to be representative for the demands placed on the driver during his changed role, the scenarios were divided in two categories; so called 'hazardous' and 'critical' scenarios. The hazardous scenarios encompass a need for driver's supervision due to reduced

Table 1
Defined ‘hazardous’ and ‘critical’ scenarios.

‘Hazardous’ scenarios (requesting supervision from the driver)	
1a “Complex Road” Attention-need induced by infrastructure	Attention is needed because the ego-vehicle approaches an on- and off-ramp where vehicles can simultaneously enter and exit the main road
2a “Vehicle Passing” Attention-need induced by traffic-participant	While the ego-vehicle is driving at the left lane, extra attention is needed because a vehicle illegally overtakes at the right
3a “Speed Oscillation” Attention-need induced by system design	When speed transgresses 50 km/h attention is needed because the system terminates if follow-speed remains above this speed (i.e. the system’s speed threshold)
‘Critical’ scenarios (requiring driver’s intervention)	
1b “Emergency Brake” Intervention-need induced by system design	The target-vehicle performs an emergency brake unexpectedly with deceleration rates exceeding system boundaries. Without intervention a collision is inevitable
2b “Merge-out” Intervention-need induced by system design	The target-vehicle leaves the lane suddenly. Without a new target-vehicle, the system requests the driver to take back control and stops automation. Danger avoidance involves preventing the vehicle from leaving the lane
3b “Close Cut-in” Intervention-need induced by combination of behaviour traffic-participant & system design	A neighbouring vehicle cuts in and brakes strongly in an attempt to take an exit. Since system boundaries for minimum required follow-distance are violated, a collision would be inevitable without intervention (see also Fig. 3)

system’s reliability. For example, the ego-vehicle approaches a combined on- and off-ramp with other vehicles simultaneously entering and exiting the main road with short follow distances. The ‘critical’ scenarios required intervention while the automation was terminated. Within two critical scenarios, i.e. “1b: Emergency Brake” and “3b: Close Cut-in”, intervention was necessary to avoid an accident. Apart from differentiation in the two main driver’s tasks (i.e. supervision and intervention), in both categories the scenarios are also differentiated with regard to the elements in the traffic-road situation that cause increased demand for driver’s support. Examples are demand induced by changes in *infrastructure*, behaviour of other *road users*, or *system-design choices* (see Table 1). With these elements the scenarios are representative for the dynamic character of road-traffic situations driving automation will encounter. People might criticise that due to improved system design (e.g. earlier recognition of new target when a vehicle cuts-in) intervention might not always be needed within the scenarios we have been using. However, as longs as the driver remains responsible and is required to act as a back-up in case automation fails, scenarios which require take-over in unexpected situations remain relevant for our research on support for intervention. (Readers who are interested in a more detailed description of the scenarios and explanation of their applicability for assessment of driver’s support when driving automatically, are advised to read: Van den Beukel et al. (2015)) The automation was modelled in our driving simulator environment as a congestion assistance, called “Traffic Jam Assist” (TJA). The modelled system is comparable to systems currently under development; cf. Naujoks, Purucker, Neukum, Wolter, and Steiger (2015). Fig. 3 provides a schematic overview with system boundaries. The own vehicle is called “ego-vehicle” and the vehicle in front which is being followed is called the “target vehicle”. The TJA-system basically keeps the ego-vehicle in the centre of the lane at a pre-set velocity and keeps a safe distance to the target-vehicle in front. To keep its lateral position detection of road-lines and a target-vehicle are required. For safety reasons, systems are often restricted to drive automatically beyond a specific speed threshold. In our case, this speed threshold was set to 50 km/h. Based on our modelled system, all scenarios involved highway driving at relatively low speed with dense traffic.

In addition to support for intervention within afore mentioned scenarios (i.e. “1b: Emergency Brake” and “3b: Close Cut-in”), we focus with the four other scenarios specifically on support for supervision. The ‘hazardous’ scenarios request supervision due to: (a) changes in road complexity (scenario “1a: Complex Road”), (b) hazardous behaviour of a neighbour-vehicle (scenario “2a: Vehicle Passing”), and (c) system behaviour (scenario “3a: Speed Oscillation”). The fourth scenario is labelled “2b: Merge-out”. It is characterised by a lost target vehicle due to the vehicle in front merging out to the neighbouring lane leaving an empty lane in front of the ego-vehicle. Within this scenario the driver was requested to take over, but there was no direct danger of an accident. Furthermore, an additional condition in this scenario involved a detection task, i.e. detection of potential hazard while entering a combined on- and off-ramp.

2.3. Simulator environment

The study was undertaken in a driving simulator featuring 180° viewing angle with three projection screens, see Fig. 4. The projection provided a simulated motorway environment with traffic-behaviour in line with the scenarios described. The simulator was not moving-based. The absence of forces that simulate a vehicle’s dynamic behaviour is not regarded as influencing our test, because the scenarios involve low-speed motorway cruising without tight curves. Participants were seated in a mocked-up vehicle equipped with common automobile control interfaces, including a physical steering wheel, physical gas and brake pedals and an automatic gearbox. Mirrors and a speed indicator were projected onto the outside screen. Instructions how to control the driving simulator was provided at the beginning of the experiment and included time to practice. Other vehicles drove in front, aside and behind the simulated vehicle. All vehicles drove with time headways

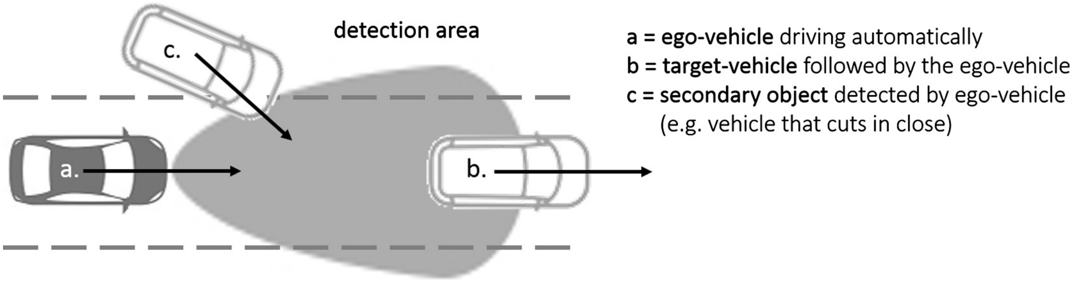


Fig. 3. Schematic overview with main system boundaries of TJA. These boundaries consist of: (i) recognition of road lines, (ii) recognition of a target-vehicle (b), (iii) driving on motorways and (iv) driving below a speed threshold, for instance 50 km/h.

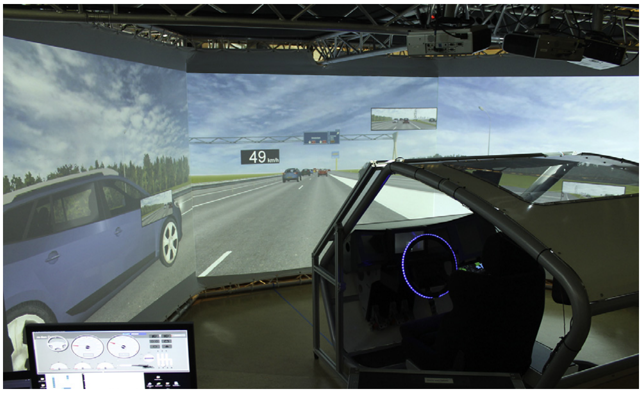


Fig. 4. Driver simulator environment.

between 1 and 1.5 s at about 50 km/h, as to simulate traffic congestion. The position of the neighbouring vehicles was identical within each trial of a specific scenario to ensure that every participant carried out the situations under the same circumstances. The controls allowed participants to intervene any time and to take back full control when necessary. A display behind the steering wheel showed the graphical interface for concept B. Furthermore, the edges of the windscreen and side windows were equipped with LED-strips to provide the illumination signals for concept C. The behaviour of the concepts was synchronised with the traffic behaviour by pre-defined scripts. The simulated environment allowed the experimenter to start the traffic scenarios and corresponding scripts simultaneously.

2.4. Experimental design & procedure

A within-subject experimental design was implemented. All participants completed three sessions (one with each concept) and each session consisted of four different scenarios. The scenarios were selected from the six available scenarios explained in Section 2.2. The different scenarios were assigned to ensure that participants encountered in each session (i.e. with each concept) two scenarios which required intervention and two scenarios which required supervision. Furthermore, the scenarios were divided over the sessions per participant to ensure: (a) participants experienced each scenario twice but always with another concept, and (b) the combinations of a concept with a specific scenario, accounted for order-effects. Each subject participated in 12 trials (3 concept-sessions $*$ (2 intervention + 2 supervision scenarios) = 12 trials). The test set-up planned for at least 36 participants, providing data from 432 (36 $*$ 12) trials, and subsequently 24 (432/(3 concepts $*$ 6 scenarios) = 24) data-sets were planned per concept tested within a specific scenario. A trial lasted between 2 and 3 min. Time between trials was needed for taking test queries. In total an experiment took per participant a maximum of one hour and fifteen minutes.

Participants were instructed that they were the responsible driver of a vehicle that was driving automatically within the following boundary conditions: Travelling on motorways with speed until 50 km/h, during which a target vehicle and road lines were recognised. The automation was introduced in terms of a Traffic Jam Assist system and it was explained that situations could happen which require extra attention (like approaching an interchange) or require the driver to take over control (e.g. when road lines are not being recognised). To help the driver with his/her responsibility, it was also explained that the system would give signals (i.e. stimuli provided by the concepts) to either ask for attention or urge for intervention. Participants were not familiarised with the signals before taking part in the trials. To add on realistic circumstances, a tablet mounted at the centre console enabled participants to view news from internet-sites or select music from a stored database.

Furthermore, participants were explained that they received full vehicle control as soon as they used any of the regular controls (i.e. steering wheel or pedals). A test run was performed by the participants to exercise on operating the simulator, including braking and lane-changing. The participants were encouraged to use the tablet as long as the driving situation allowed to do so, according to their own judgement. The instruction intended to provide trust in the automation and it was advised to only intervene when intervention seems indispensable to avoid an accident.

2.5. Selected measurements

The concepts were tested for intervention-support during the critical scenarios that required accident avoidance (i.e. "1b" and "3b"). Accident avoidance was measured with Time-to-Collision (TTC) and Accident Avoidance Ranking (AAR). Similarly as described in Van den Beukel et al. (2015), AAR accounted for three-fold possibilities of intervention: (i) intervention failed and an accident occurred, (ii) intervention was successful and consisted of braking while staying within the original lane, or (iii) intervention was successful and based on swerving out without causing (as yet) an accident. One scenario ("2b: Merge-out") considered detection of changes in system state due to a failing target-vehicle (A previous target merged out leaving an empty lane in front). Subsequently, the participant was requested to take over driving. Within this scenario performance was evaluated by observing take-over behaviour and measuring reaction times until participants took over control. Supervision is considered successful if drivers remain situation aware. In our case, situational awareness refers to both traffic situation and automation state, for example system's recognition of a target vehicle. A wide-accepted definition of Situation Awareness (SA) is provided by Endsley (1995) and describes SA in terms of three levels: "(1) the *perception* of elements in the environment within a volume of time and space, (2) the *comprehension* of their meaning, and (3) the *projection* of their status in the near future". Correct SA was measured for each concept in every scenario based on a specific probe-taking technique, i.e.: SAGAT (Charlton, 2002; De Winter et al., 2014; Endsley, Selcon, Hardiman, & Croft, 1998). SAGAT is based on freezes in the driving simulation and then probes are taken that are assumed to be representative for SA as they resemble the three levels of SA. For example: After the simulation froze, it was asked: "What type of vehicle is aside you?" (*perception*), respectively; "What caused the system to request your attention?" (*comprehension*) and "if the simulation continues after the "freeze", what situation would most likely require your intervention?" (*projection*). We also examined Workload, Arousal, Spare Attention and Received Information, based on self-assessment. These four aspects represent mental efforts and were measured with queries,[4] like: "How much information have you gained about the situation? Have you received and understood a great deal of knowledge (High) or very little (Low)?" Participants were then asked for their self-assessment by answering on a 7-point scale ranging from 1 (low) to 7 (high). Furthermore, to gain insight in participants' subjective appreciation of the concepts, we measured concept acceptance after each session with a 'Van der Laan'-questionnaire (Van der Laan, Heino, & De Waard, 1997) that measures two factors: Perceived Usefulness and Satisfaction. (In comparison to the original 5-point scale, we adapted the Van der Laan scale to contain a scale with 7-points in order to provide consistency with afore explained 7-points scale for measuring mental efforts.) At the end – after the experiment was completed and participants had experienced all three concepts – participants were also asked for their preferred concept. Then, participants expressed their preference by assigning 15 points over all three concepts. This allowed participants to express a balanced opinion. They could for example express indifferent scores by assigning 5 points equally to each concept. Or, they could express scores to two almost equally superior concepts by a 0/8/7 divide. Any division of scores between concepts was allowed as long as they add on to a total of 15.

3. Results

3.1. Support for intervention

With regard to support for intervention, our concepts have been tested with scenarios "1b: Emergency Brake" and "3b: Close Cut-in". (The results for intervention-support are a summary of a previous study conducted by the authors (Van den Beukel & Van der Voort, 2016).) The two scenarios both provided a necessity to intervene and avoid an accident. Because participants experienced more scenarios and the scenarios were offered in randomised order, they were unprepared when and in what circumstances an intervention would be required. The expectation was that concept C would provide strongest support due to directing visual attention outside the vehicle towards the locus of a critical event. Besides, the tactile stimulus was expected to provide a strong affordance for appropriate action (i.e. braking). Because concept B provides graphical information with a depiction of system state, this concept was expected to provide especially support for supervision and to a minor extent for intervention. Nonetheless, the results did not show evidence that either B or C provided additional support for intervention of any substantial effect compared to the baseline (A). On the contrary, it was hypothesised that concept C might have influenced intervention counter-productively.

With regard to support for intervention, the results referred to Accident Avoidance, gained Situational Awareness and Cognitive Performance. Most accidents occurred with concept C (10 out 49 occasions = 20.4%). Concept A denoted accidents in 7 of 48 occasions (14.6%) and participants encountered least accidents with B (6 out of 47 = 12.8%), see Fig. 5. Highest scores for accident occurrence with concept C were against expectations. A concept-ranking was also made to provide more

---

[4] The queries were taken from a standardised SART questionnaire (Charlton, 2002).

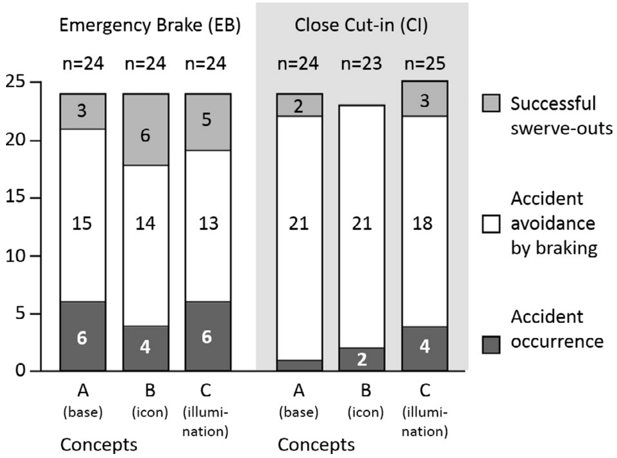


Fig. 5. Numbers with failed intervention (i.e. accident occurrence, shown with the bottom bars) and successful intervention, either by braking (middle bars) or by swerving out (top bars).

insight of how well participants avoided accidents. In addition to the numbers provided in Fig. 5, this ranking also reflected performance *differences* in accident avoidance by braking, like: just in time, with small or with large time-to-collision (TTC). These differences were based on TTC-values. In other words: The ranking accounted for the following: (i) failed intervention (worst ranking); (ii) successful accident avoidance by braking and staying within the original lane (intermediate ranking based on TTC-values), or (iii) successful intervention based on swerving out without causing (as yet) an accident (best ranking). For further explanation on the calculation, see Van den Beukel and Van der Voort (2016). When results from both scenarios are taken together, the test provided 138 possible ranks. Concept B was best ranked with a mean rank of 63.0, concept C scored a mean rank of 70.3 and concept A scored a mean rank of 75.4. These scores do not reveal statistically significant effects ($\chi^2(2, N = 138) = 2.22$, p-value = 0.33). Based on accident occurrence, scenario 1b was most dangerous (see Fig. 5). Differences in ranking scores are larger within this scenario, but again without showing statistically significant results. Nonetheless, scores show the same pattern: B was ranked best, C mediocre and concept A worst. Despite lack of statistically significant results, the observation that B scores best and C in some occasions worse (i.e. for accident occurrence) was against expectations and therefore raises the concern that C might have been counter-productive.

When considering these results we need to remember that concept C did not only provide a visual-acoustic stimulus but also a vibration at the front-end of the seat-pan at the moment intervention was required. Measurement of psychological aspects revealed that – compared to the baseline concept – the additional stimuli provided by concept C does increase arousal and at the same time make the use-situation more demanding and reduces spare attention. It was hypothesised that some characteristics of concept C (i.e. attracting attention without a specific instruction) had made people more awaiting with regard to what happened and whether the system might intervene or take some sort of control by itself. This could explain why concept C had higher (worse) Accident Avoidance Ranks than concept B even though concept B required reading of instructions (i.e. icon with text). In this respect the better performance (though without significant differences) for concept B could be the result from concept C being less supportive instead of B being more supportive. Although this could have influenced concept scores, most likely the combination of stimuli provided with C has negatively influenced performance, due to making situations appear more dangerous. Indeed, seventeen (of 37) participants mentioned explicitly that they perceived the illumination-concept (C) to be attracting attention and at the same time being ‘too much’, causing annoyance. This is a plausible explanation since mismatched perceived levels of urgency (too low, but certainly also too high) has shown in previous studies to influence performance negatively (Baldwin & Lewis, 2014; Chapanis, 1994). Moreover, the combination of stimuli, including tactile feedback at the moment intervention was necessary, has probably caused a startle-response for some participants (Baldwin & May, 2011).

Concept C was composed to not include vibration during ‘soft warnings’ (i.e. for supervisory tasks). It could be that supervision and intervention benefit from different concept-compositions. Therefore this paper continues to examine the defined concept in additional scenarios, including those demanding supervisory task-performance. The results will be compared with above explained results for intervention and allow a more in-depth evaluation what concept types are beneficial for the driver’s changing role, i.e. supervision with (now-and-then) intervention.

3.2. Support for taking back control

The necessity to take-back control in scenario 2b was influenced by system-design, i.e. the inability to drive automatically without a recognised target-vehicle. Although participants had received explanation on the system’s boundary conditions, this scenario was one of the most difficult to understand. Because there was no training on exploring system boundaries.

Besides, it is always more difficult to deduct information based on something that is missing (i.e. a target-vehicle) compared to when elements are added in the traffic situation (e.g. a vehicle cutting in). Regarding the understanding of the need to retrieve control: When using concept A participants understood this in 39% of occasions, against 56% and 55% of occasions for respectively concepts B and C. However these differences do not denote statistically significant differences and there was no effect of Concept on Taking back Control: $\chi^2(2, N = 68) = 1.62$, p-value = 0.44.) Although these results are in favour of concepts B and C, review of Reaction Time (RT) provides a different picture. Then, concept A scores best. RT was calculated based on the differences in time between the moment a take-over-request was raised (both acoustically as visually) and the moment when the participant used the accelerator. This was done while after the ego-vehicle lost its target, longitudinal control (i.e. ACC) was terminated and using the accelerator was taken as indication for participants consciously noticing its termination and performed taking over of control. Again the effect of Concept on Reaction Time was not statistically significant: $F(2, 31) = 1.24$, p-value = 0.30. However, reaction time was best (fastest) with concept A (Mean = 4.32; SD = 1.43) and slowest with C (mean = 5.31; SD = 1.46). Concept B scored mean = 4.88 and SD = 1.33. Interestingly, the concepts that related to most occasions with successful retrieving control (i.e. B and C) denoted longer reaction times compared to A. This can be explained by the additional information or cues which B and C provided, which needed time for interpretation and therefore denoted high RT. But this additional information could have helped more participants to understand (and subsequently decide) to take-over vehicle control. Results show a trend RT for A is better compared to C: p-value = 0.059 (two-tailed) (see Fig. 6 and Table 2).

3.3. Support for supervision: hazard detection

The detection task allows to compare only the additional visual stimuli provided by concepts B and C. This is because in the situation where the detection task was being raised, only visual stimuli were being transgressed. The situation was the merge of two highway sections. At this occasion the sides of the windscreen lit op with concept C and concept B displayed a graph with side-warning, see Fig. 7. Both concepts did not raise an acoustic alert in this situation, nor tactile feedback. Therefore comparison of detection in this situation allowed if any of the visual feedback alone (icon or illumination) intrinsically raised more attention. Table 3 shows that the statistically significant differences are in favour of illumination (concept C): With the icon-feedback (concept B) participants successfully detected the alert in 40.9% of occasions, and with illumination detection was significantly higher with 82.6% of occasions ($\chi^2(1, N = 45) = 8.32$, p-value = 0.004, two-tailed).

3.4. Support for supervision: Situation Awareness (SA)

This section provides results of how situation aware participants were while being supported by a concept in each of the scenarios demanding a supervisory-role from the driver, i.e.: “1a: Complex Road”; “2a: Vehicle Passing” and “3a: Speed Oscillation”.

Scenario 1a, approaching a combined on- and off-ramp. Within this scenario the attention-need was induced by infrastructure, i.e.: an increasing road complexity with traffic from a combined on- and off-ramp dominated the situation. MANOVA did not show a significant influence of Concept on SA-*Perception* ($\chi^2(2, N = 72) = 4.06$, p-value = 0.132, two-tailed), neither on SA-*Understanding* $\chi^2(2, N = 72) = 3.67$, p-value = 0.160, two-tailed). Nonetheless, separate ANOVAs showed that SA-*Perception* was with concept C significant better than with B: $\chi^2(1, N = 48) = 4.36$, p-value = 0.019 (one-tailed). SAGAT-perception also shows a trend to favour C over A: $\chi^2(1, N = 48) = 3.20$, p-value = 0.037 (one-tailed). The results show that within this scenario requiring attention due to changes in road infrastructure, perception of elements important to gain Situation Awareness (SA), was best with the illumination-concept (C) and worst with the icon-concept (B). *Understanding* of the elements important for Situation Awareness appears equally good with A and C and worse with concept B. See Fig. 8a. However, no significant difference were found between concept B and either concept A or C. It is likely that the visually dominated changes of road complexity seen in this scenario, comprise circumstances where the directional illumination (concept C) demonstrate its full potential.

Scenario 2a, vehicle passing, offered a situation of a neighbouring vehicle at the right lane while overtaking and therewith created a potentially dangerous situation. The situation was designed to demonstrate a potential cut-in from the right. Hence, the attention-need was induced by the behaviour of other road users. The perception of a vehicle being in the blind spot was with all three concepts high: between 92% and 100%. Between 90% and 100% of participants correctly understood the risk of a possible cut-in. However, only between 50% and 58% of participants also correctly anticipated future state that this could cause a necessity for taking-over. The scores between concepts were indifferent and therefore do not provide in this situation insight if one of the concepts provide more support for supervision than the other.

Scenario 3a, speed oscillation, comprised a difficult to understand situation. Complex system behaviour, i.e. traffic speed oscillating around a defined speed threshold, raised a soft warning to prepare drivers for potential take-over in case the system boundaries would be transgressed. The difficulty of this scenario is reflected on the relatively low scores for SA-Understanding, see Fig. 8b: It varies between 16% correct understanding with concept B and 39% correct understanding with concept C. These difference reveal a trend that correct understanding in this scenario is higher with concept C than concept B: $\chi^2(1, N = 48) = 2.64$, p-value = 0.052 (one-tailed). Anticipation of the future-state (i.e. probability to take-over control) was low (about 30% correct anticipation) and did not demonstrate differences between concepts.

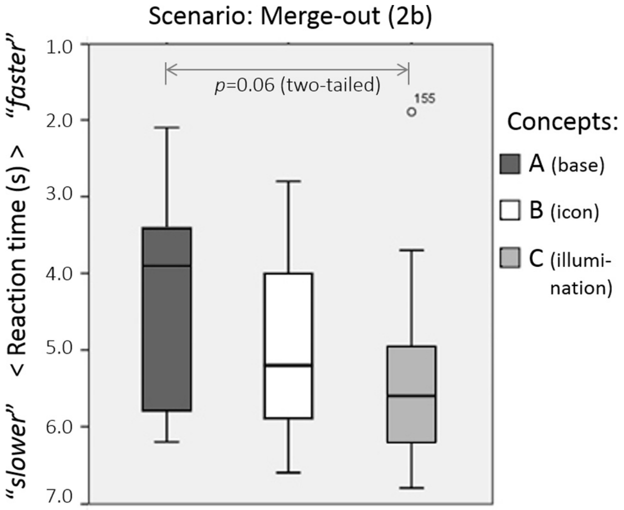


Fig. 6. Reaction time to take back control – scenario: merge-out target-vehicle (2b).

Table 2
Number of successful performance to take back control. Scenario: Merge-out target-vehicle (2b).

Taking back control – scenario: Merge-out (2b)			
Concepts	Successful	Failed	*N*
A (base)	9 (39%)	14	23
B (icon)	14 (56%)	11	25
C (illumination)	11 (55%)	9	20
Total	34 (50%)	34	68

3.5. Support for supervision: cognitive performance

Cognitive performance was measured in the same scenarios reviewed in the previous section. The assessment comprised the following: Workload, Arousal, Spare Attention and Received information - all based on self-assessment using standardised questionnaires (Endsley et al., 1998). One-way ANOVAs showed that there were no significant effects of Concept on any of the cognitive aspects within the scenarios. Nonetheless, some results are noteworthy.

In scenario 1a ("Complex Road"), only a mediocre influence of Concept on Received Information was found: F (2, 69) = 1.38, p-value = 0.26. Results with regard to Received Information denote largest differences between Concept A (Mean = 4.75, SD = 1.39) and Concept C (Mean = 5.40, SD = 1.28), indicating a trend on participants' perception to receive more information in scenario 1a with concept C than with A: p-value = 0.051 (one-tailed). Scenario 2a did not denote any notable differences.

Within scenario 3a ("Speed oscillation") we observe some differences between received information and arousal. Participants perceived to receive more information with concept B (Mean = 5.06, SD = 1.43) compared to A (Mean = 4.40, SD = 1.70) or C (Mean = 4.42, SD = 1.18). Probably this is due to concept B being most specific within this scenario. However, Perceived Information does not show significant differences between the concepts: Between A and B p-value = 0.12 and between C and B p-value = 0.13. Furthermore, in this scenario participants perceived themselves less alert with concept A (Mean = 4.23, SD = 1.56) than with B (Mean = 5.06, SD = 1.55) or C (Mean = 4.75, SD = 1.73). The situation was difficult to understand without interface support. Presumably, most participants perceived it as "not much going on" (as some participants also stated) and maybe this has caused ignorance.

3.6. Concept acceptance

Table 4 shows the mean scores of subjective rating for Perceived Usefulness and Satisfaction (Van der Laan et al., 1997) based on scores ranging[5] from: 1 = absolutely not useful, not satisfying, to; 7 = extremely useful, satisfying. Based on one-way ANOVAs, no significant effect was found of Concept on Perceived Usefulness (F(2, 106) = 0.37, p = 0.69) and neither on Satisfaction (F(2, 106) = 0.13, p = 0.88). Of course, baseline subjective ratings differ per person. Therefore a within subject comparison of

---

[5] Due to reasons explained in Section 2.5 the original 5-points scale was adapted to a 7-points scale.

Fig. 7. Pictures showing concept B (left) and concept C (right) during the hazard alert for an approaching vehicle in a highway-merge situation. Concept B depicts an icon behind the steering wheel, and concept C uses illumination of the pillar aside the windscreen. Both alerts were only visual.

Table 3
Number of successful performance detection-task.

Concept	Detection		Total
	Successful	Failed	
B (icon)	9 (41%)	13	22
C (illumination)	19 (83%)	4	23

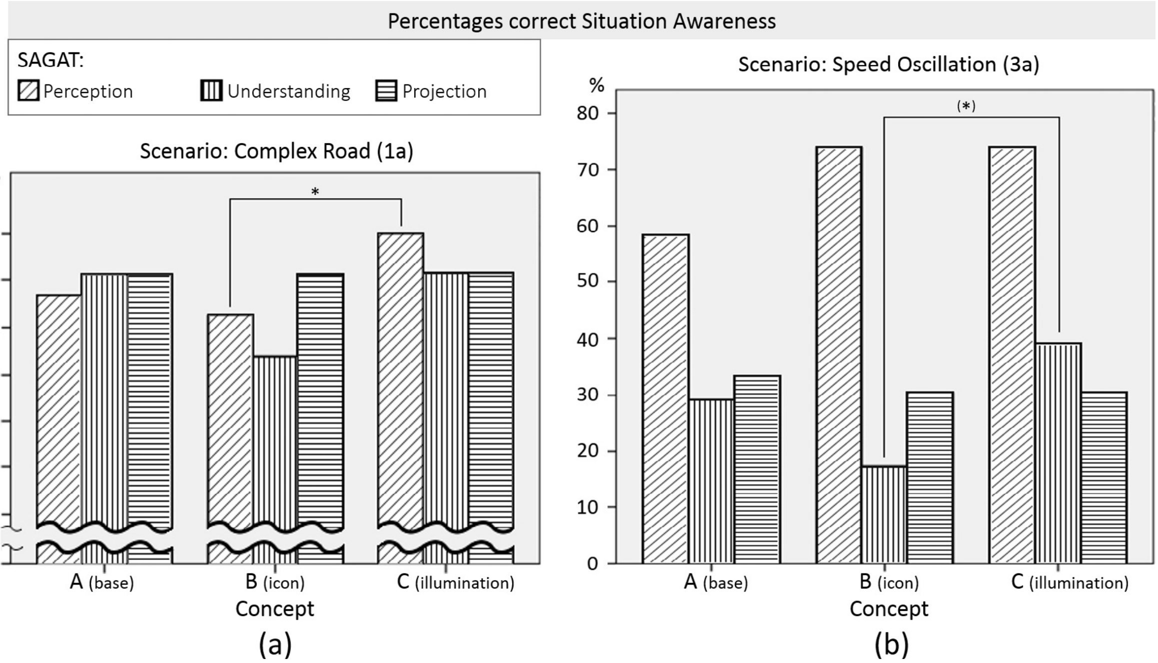


Fig. 8. Percentages correct Situation Awareness (based on SAGAT) per SA-level (SA-Perception, SA-Understanding and SA-Projection) and per concept. Part (a) shows results within scenario Complex Road (1a) and (b) within scenario Speed Oscillation (3a). Results marked with "∗" denote statistically significant differences ($p < 0.05$). Differences between $0.05 < p < 0.10$ indicate a trend and are marked with "(∗)".

differences in ratings is provided in Table 5. The table compares scores for concepts B and C to concept A – which scores were, as a baseline concept, set to zero. In other words: Scores for B and C in Table 5 are taken within subject relative to concept A. Then, the largest difference between A (base) and C (illumination) show results in favour of C with respect to perceived usefulness: $t(35) = 1.174$, $p = 0.249$ (two-tailed). However, again the results do not denote significant effects of Concept (B or C) on Perceived Usefulness, neither on Satisfaction.

Table 4
Mean scores for Perceived Usefulness and Satisfaction (VanderLaan).

Concept-scores for concept acceptance						
Concepts	Perceived Usefulness (VanderLaan)			Satisfaction (VanderLaan)		
	N	Mean	Std. Dev.	*N*	Mean	Std. Dev.
A (base)	36	5.70	1.14	36	4.84	1.24
B (icon)	36	5.77	1.23	36	4.96	1.32
C (illumination)	37	5.92	1.05	37	4.99	1.32
Total	109	5.79	1.14	109	4.93	1.28

Scores range from 1 = absolutely not useful, not satisfying; to 7 = extremely useful, satisfying.

Table 5
Relative concept-scores, compared to baseline-concept (A).

Concepts	*N*	Perceived Usefulness (VanderLaan)			Satisfaction (VanderLaan)		
		Mean	Std. Dev.	*p*-value cf. A (base)	Mean	Std. Dev.	*p*-value cf. A (base)
B (icon)	35	0.059	1.447	$p = 0.81$	0.093	1.301	$p = 0.68$
C (illumination)	36	0.194	0.994	$p = 0.25$	0.097	1.458	$p = 0.69$

Scores are within-subject comparisons with scores for Perceived Usefulness and Satisfaction for concept B, and concept C, compared to concept A. For example, the mean Perceived Usefulness score for B (0.059) is the average difference of all participants between scores for concepts B and A on Perceived Usefulness.

Table 6
Participants' subjective concept rating.

Concepts	*(a)* **Preference:** Numbers of participants' selected concept as		*(b)* **Points** (SUM of 15 available) assigned to concepts		
	"preferred"	"poorest"	Mean	Std. Dev.	% of 15 points (%)
A (base)	6 (18%)	24 (70%)	*3.47*$^{*,\#}$	3.79	23
B (icon)	9 (26%)	7 (21%)	*5.56*$^{\#}$	3.46	37
C (illumination)	19 (56%)	3 (9%)	*5.97**	3.09	40

Participants' subjective concept rating expressed with (a) concept-preference and (b) assigned points, i.e.: Participants were asked to divide a total of 15 points over the three concepts. Statistically significant differences are typed in italic font and indicated with * or $^{\#}$.

Table 6 shows participants' subjective concept rating. This ratings was expressed with (a) participants' selection of their preferred concept, and (b) the participants were asked to divide 15 points over the three concepts. The results show that a majority (56% of respondents) preferred Concept C and concept A was most rejected (by 70% of respondents). The low scores of concept A are according to expectations as it was selected to only provide audible signals, i.e. least support. With regard to the assigned points, see (b) in Table 6, the differences between A and C are (very) significant: $t(70) = 3.068$, $p < 0.025$. Differences between A and B are significant: $t(70) = 2.455$, $p < 0.025$. Differences between B and C are not statistically significant ($p = 0.610$). Basically, the subjective ratings shows scores in favour of concept C but the differences with B are small. Based on participants' subjective opinion concepts B and C are both accepted and concept A is rejected.

3.7. Participants' perception of concepts

At the end of the test each participant was conducted to an open interview. During this interview the experimenter first asked what different stimuli they could recite. This was to activate participant's memory of the concepts he/she encountered during the trials. Then the experimenter summarised shortly in what order the participant had experienced which concept and asked for their spontaneous opinions, e.g. likes and dislikes, irritations and surprises or any other opinions about the concepts. The interview was closed with asking: (a) participant's preference, and (b); to divide 15 points over the three concepts – as explained in the previous section. This section presents an overview with the participants' spontaneous opinions expressed during the interviews.

Table 7 summarises the main findings from the interviews. Concept C evoked the most ambivalent opinions. On the one hand a vast majority (28 participants) considered concept C to be attracting attention (considered a positive asset) and at the same time 21 participants explicitly mentioned that they perceived some aspects *how* concept C attracts attention unpleasant. Moreover, in five occasions (when experiencing concept C's hard warning) participants reported during the experiment that they had a fright. We observed different aspects causing participants' negative opinion: 6 participants mentioned explicitly that they found the vibrotactile stimulus in the seat-pan to be irritating. 7 persons found the illumination – especially the

Table 7
Number of participants who associated a concept (A, B or C) with a particular opinion.

Opinion	A	B	C	Example
Concept attracts attention in a positive manner	11	2	28	Signals explain what needs attention
Concept attracts attention in a negative manner	1	3	21	Signals are too harsh
Concept is informative	–	15	6	Signals confirm correct system function
Concept is not informative	5	9	3	Information was not recognised during trial

Measurements are based on spontaneous answers during an open interview at the end of each experiment.

red illumination – annoying. Two people perceived the illumination to be unnecessary and only providing a distraction. Furthermore, 3 participants rejected concept C due to lack of explanation why an alert was raised. Seventeen participants found concept C to be both attention attracting *and* annoying at the same time. Most people recognised that the stimuli differed between soft and hard-warnings and that their rejection how attention was attracted especially considered the hard warnings. This could explain why concept C was preferred and at the same time conceived to attract attention in a negative manner. Although the vibrotactile stimulus of concept C mostly received rejection, 3 participants explicitly mentioned the haptic feedback to be positive. They made remarks like: “It makes you alert right away”. “Building up trust” (in automation - induced by concept C’s feedback) was also mentioned three times.

The illumination around the steering wheel received ambivalent reactions as well. 5 people generally appreciated the illumination and 9 people rejected it. Some people recognised the steering wheel illumination as confirmation of system activity, 1 person perceived the wheel illumination more quietly than windshield-illumination. However, 2 persons perceived it an unpleasant feature or a distraction. According to 2 people it does not offer added value and 1 person mentioned that it brings you in a position to look away from the road. When reviewing these results, we need to keep in mind the limitation that participants did not operate the automation themselves. With all trials participants drove directly automatically which was also explained beforehand. There was no need to activate or deactivate the automation during the trials. Therefore the need to check system activity during the trials was relatively low and possible added value to verify with the steering wheel illumination whether the system is active, could not be experienced.

For being informative, especially concept B was favoured (15 times). The dominating reason was that concept B provided explanation on cause for system-status and -change. Despite the reduced need to check system activity, 6 participants explicitly mentioned that concept B allowed anytime a short check on system status. Some people also valued that the concept provided an explicit instruction to the driver, like “Take over now” or “Attention complex road”. Only 2 participants mentioned their appreciation for the way B attracted attention, explaining that it was the location they were familiar with. In contrast, 17 participants rejected the way concept B attracted attention or considered the information not informative. The main reasons were: Necessity to look away from road (2); Information was unclear (1); Information was not being noticed (6) or obstructed by steering wheel (13 times). Here, it needs to be mentioned that the instrument cluster behind the steering wheel *only* depicted information on automation status. With a more commonly applied set of information, the demand to review information in the cluster would be higher and this might raise appreciation about concept B’s location at the same place of other instruments.

Eleven people were positive about the attention attracting feature of concept A. This confirms that sound is an important base feature to attract attention. The remarks of some participants seem to reflect that they found the additional features of B or C not productive and they had the opinion that a sound alert would be enough. Even more, some people expressed that the additional features would only be a distraction and for that reason favoured the baseline concept (A). Others (5) were critical against A: “The sound causes a fright” (1), and “The sound does not explain why an alert is raised” (2) were mentioned.

Independent of the concepts, 17 people mentioned explicitly that they considered usage of a tablet to be too distracting and dangerous. (During our test participants were asked to browse the internet and select sound tracks from a stored database.) People remarked that they only performed this secondary task because they were asked to do so. Observation of the task showed that it was not performed in an effective way: Most people browsed a bit on news-sites, without actually reading.

Five participants were afraid that the illumination provided with concept C would cause glare for themselves or other road users. Influences of weather and light conditions on the ‘readability’ of concept C’s illumination signals are an important topic for further research. One participant named the yellow illumination field with the wrong colour, i.e. “green”. This person denied to have dichromatism. An explanation could be that the yellow illumination against the blue coloured sky in the simulator made the illumination appear green. Of course, it is important that the level of urgency conveyed with the illumination-signals is not being influenced by light conditions.

4. Discussion

We tested three concepts: Concept A which provided only acoustic warning and served as baseline; Concept B which required interpretation of graphical information on system status, displayed as icons behind the steering wheel; and Concept C which combined an acoustic-tactile stimulus with ‘directional’ illumination. See Section 1.4 for a detailed description of the concepts. The aim of the test was to explore what interface-aspects provide on the one hand support for supervision and on the other hand support for intervention. The concepts provided stimuli which are differentiated between soft and hard warnings. The soft-warnings demanded attention for potential hazardous situations. The hard-warnings required intervention to take over control. Concept C was designed with a large difference between both warning types: Its hard-warning offered not only colour coding (consisting of red illumination of complete lower windscreen-band), but also included an additional tactile stimulus (vibration of the front seat pan). During concept creation important considerations were to inform drivers about system-status and to support them in returning back in the control-loop, in order to reduce automation-induced OOTL-performance problems – as explained in the introduction.

4.1. Performance expectations and main results

The expectation was that especially the acoustic-tactile concept with 'directional' illumination (concept C) would provide best support for drivers in their changed role. This expectation was based on the concept's ability to attract and direct attention towards the locus outside the vehicle that needs attention. It was also hypothesised that concept C – in contrast to B – would urge the driver to notice why attention was needed within the traffic scenery itself. Therewith, C was expected to induce stronger driver-involvement and made him/her take part in the control-loop, whereas concept B was expected to require reasoning-time until the driver had understood where attention is needed. Therefore, faster reaction times were expected with concept C compared to B. Main findings are summarised in Fig. 9. In short, during 2 (out of 3) scenarios which require supervisory support, the illumination-concept (C) showed significant or close to significant better Situational Awareness (SA) than the icon-concept (B). Moreover, with the illumination-concept (C) the detection-task was performed significantly better than with the icon-concept (B). Against expectations however, neither the illumination-concept (C) nor the icon-concept (B) demonstrated support for intervention. Presumably, support for supervision and intervention benefits from different interface-features.

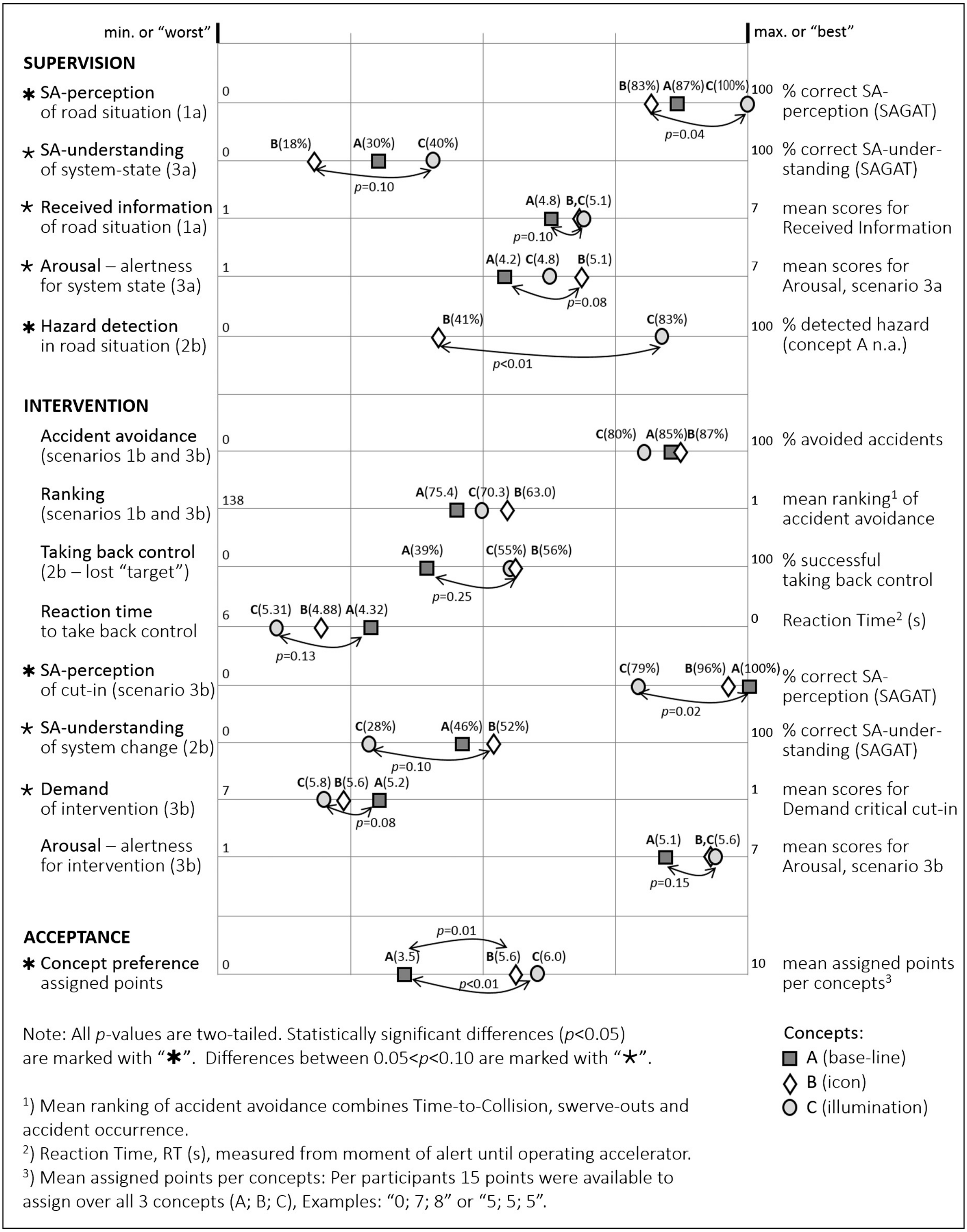
We will now discuss what these results mean for recommending interface-features to provide the desired support for both supervision and intervention.

4.2. Driver-interface support for supervision

The results suggest that guiding attention outside the vehicle with the illumination-concept (C) helped drivers more to gain SA than displaying information in the vehicle with the icon-concept (B). The latter required interpretation what outside traffic-aspects the displayed information should be connected with. The presumption is that icons used for feedback on system-state leave the driver to remain a rather passive 'observer', whereas the directional illumination provides a stronger cue to return in the control-loop. We also expected that the required interpretation *why* an alert was raised with the illumination-concept (C) would be a disadvantage. The icon-feedback on the other hand, required interpretation *where* traffic situations outside the vehicle needed attention, but was expected to convey more easily complex matters (for instance: changes in system-state induced by system-design). Nonetheless, within the speed oscillation scenario ("3a": a scenario in which an alert was raised because of potentially changing system-state due to system's speed-threshold) the icon-concept (B) did not demonstrate raised performance. Maybe this is due to the icon-feedback not being observed behind the steering wheel, or due to misunderstanding of the provided icon-feedback. It could also be that support *where* attention is needed, compensates for reduced support explaining *why* attention is needed. These speculations leave unimpeded, that the illumination-concept (C) showed in the scenarios requesting supervision, superior levels of SA. Moreover, the illumination-concept (C) showed to be also superior with respect to the detection-task. Indeed, several studies confirm that driver's performance improves (e.g. better reaction times and accuracy) if attention is guided towards relevant locations, because operator's control improves by the information and direction of danger provided (Lee & Chan, 2007; Liu & Jhuang, 2012). Furthermore, the illumination-concept (C) received higher acceptance than the icon-concept (B). Overall, the results within this study showed a favourable relation between the illumination-concept (C) and support for supervision.

4.3. Driver-interface support for intervention

Despite the promising results of the illumination-concept (C) for supervisory support, neither the illumination-concept (C), nor the icon-concept (B) revealed to be of distinctive support for intervention. On the contrary – though without statistically significance –, the illumination-concept (C) showed intervention-performance poorer than the baseline concept. Therefore the assumption was that the illumination-concept (C) had some counter-productive effects. Cause for counter-productive effects on intervention-support could be inappropriate levels of urgency, while too high levels of perceived urgency cause annoyance (Chapanis, 1994; Marshall et al., 2007) and have a counter-productive influence on task-performance (Baldwin & May, 2011). Remarks made by participants actually confirm annoyance of concept C, like: "C attracts attention – to the point of being irritating". Using different terms, seventeen out of thirty-seven participants made remarks with the same tenor. In addition to the consideration that the combination of three stimuli with concept C's hard-warning could have caused irritation, we discuss below whether a specific compilation of stimuli might have caused annoyance.



Note: All p-values are two-tailed. Statistically significant differences ($p<0.05$) are marked with "✱". Differences between $0.05<p<0.10$ are marked with "⋆".

[1]) Mean ranking of accident avoidance combines Time-to-Collision, swerve-outs and accident occurrence.

[2]) Reaction Time, RT (s), measured from moment of alert until operating accelerator.

[3]) Mean assigned points per concepts: Per participants 15 points were available to assign over all 3 concepts (A; B; C), Examples: "0; 7; 8" or "5; 5; 5".

Fig. 9. Overview with concepts' main results of their support offered for supervision and intervention, including results for acceptance. The results indicate that the illumination-concept (C) offers support for supervision, but compared to the baseline (A) neither the icon-concept (B), nor the illumination-concept (C) offers substantial support for intervention.

4.3.1. Possible influence of modality on annoyance

While the differences in offered stimuli between hard- and soft-warnings were largest with the illumination-concept (C) due to the additional tactile feedback (during hard-warnings) and because the hard-warnings of concept C are associated with annoyance, one might consider that specifically the tactile feedback caused annoyance. However, an earlier study (Van den Beukel & Van der Voort, 2016a) showed that people conceived the situation with illumination to be more hazardous than without illumination in identical scenarios. These differences were based on visual aspects alone (i.e. without any tactile feedback) and the illumination areas were identical to concept C in our current study (i.e. colour, position and size of the illumination). Nonetheless, the differences were not statistically significant. With respect to possible modality-induced raised levels of annoyance, Baldwin and Lewis (2014) suggest that especially auditory signals cause deteriorating effects. However, based on Marshall et al. (2007), our particular characteristics of the audible signal are considered mediocre on both urgency and annoyance. Moreover, the sound-alert of concept C's hard-warning was identical to the one used in baseline concept (A), where it did not cause annoyance. These results suggest that within our study the sound-alert in itself did not cause negative side-effects. Although both existing research and our study do not provide strong suggestions for a causal relationship of any of these features individually, the appearance of annoyance in combination with the additional stimuli offered in concept C makes us conspicuous of some deteriorating influence of either illumination or tactile feedback.

4.3.2. Possible performance deficits due to combination of modalities

Besides presumable deteriorating influence of either illumination or tactile feedback, it could also be that the combination of different modalities (i.e. acoustic, visual and haptic within concept C) caused an interrelated conflict. Ferris, Penfold, Hameed, and Sarter (2006) revealed that invalid cross-modal cues (e.g. a driver perceives an auditory cue to come from a different location than a visual cue although both are meant to indicate the same target) lead to longer response times and reduce response accuracy. In our research the visual stimulus holds a strong spatial cue, but the tactile cue (vibration in the front end of the seat pan) was loosely linked with the direction of danger. It did not differentiate between front-left and front-right. The acoustic warning did not provide spatial cues either. According to Ferris et al. (2006) weak cross-modal links could already negatively influence performance. Therefore, weak cross-modal links of concept C could be an explanation for reduced reaction times. However, the illumination-feature of the concept holds a strong spatial link and one might argue that this would have compensated the weak cross-modal links. Furthermore, the icon-concept (B) provides even weaker cross-modal and spatial links, but is not associated with longer reaction times than the illumination-concept (C). Overall, findings from previous research on the influence from cross-modal links do not provide evidence for a specific combination of modalities within our concepts that has caused reduced intervention-performance. Nonetheless, stronger cross-modal links are important to consider for further development of desirable interface-support.

4.3.3. Possible performance deficits due to unexpectedness of stimulus

Ferris et al. (2006) discovered that response times increase due to cues in an unexpected modality. Therefore, the unexpectedness of our tactile cue within a dominantly visual task could be an explanation for the low scores on intervention-support with concept C. An extreme manifestation when a stimulus appears unexpectedly is when it causes a startle response. Measurement of startle was not foreseen within our assessment framework. However, 8 (out of 37) participants reported during the experiment that they had a fright. In five occasions this was due to the stimuli from a hard warning raised by concept C (cf. 2 times with concept B and 1 occasion with concept A). With regard to the effect of warning stimulus modality on startle, some studies suggest that startle will inhibit if the stimuli are presented in different modalities than the modality that requires attention within a specific task (Anthony & Putnam, 1985). However, several other studies found that selective attention allocated to a foreground stimulus facilitates startle regardless of stimulus modality (Bohlin, Graham, Silverstein, & Hackley, 1981; Lipp, Siddle, & Dall, 2000). This effect of selective attention on startle may not be found, if task demands are small. According to these studies, stimulus change – regardless of stimulus modality – enhances startle. Based on these findings, we suggest that the *unexpectedness* of the stimulus (in our case concept C's haptic cue) may have caused startle-responses, not necessarily the specific type of stimulus.

4.4. Limitations and long-term implications

Partially automated driving shifts the driver into a complicated role of full responsibility for driving safely without being in permanent control. The aim of this research was to propose interface-solutions, which support drivers with their changing role to supervise the automation and to intervene if required. The results show that visual illumination of relevant traffic events is especially beneficial for support of supervision. Unfortunately, the results do not allow specific recommendations on interface-aspects beneficial for intervention. During our test, the majority of participants expressed on the one hand to be positive towards the raised level of comfort induced by automation ("I like to listen to the music and relax") and at the same time rejected engagement in distracting tasks like Web browsing. On the one hand this is a positive outcome: the drivers showed to be serious about their supervisory task and are at the same time prepared to take advantage from automated driving and accept more homogenous driving (i.e. driving in follow-mode). On the other hand, this outcome is limited by test-behaviour and potential long-term implications. First of all, the participants used the system for ca. 45 min. During this time they encountered 12 trails and experienced 6 times a situation which required intervention. Therefore the test placed participants in situations where much more attention from the driver was needed than would be in real-life circumstances.

Moreover, the participants took part voluntarily out of general interest or because they considered the topic important to society. Therefore the respondents could be characterised by their consciousness of responsibility. It is unlikely that this attitude is representative for the characteristics of future generations of 'drivers' who are increasingly using automation and attention-seeking devices like smartphones. Further research is needed on long-term implications, such as habituation, behavioural adaptation and skill degradation.

4.4.1. Habituation

Habituation to a stimulus might eventually lead to a terminated response (Bouton, 2007). Especially soft warnings are in danger to be conceived as false alarms. Then, their perceived irrelevance is likely to elicit habituation. It goes without saying that ignorance of the provided signals could be dangerous, but could also cause indirect side effects like the adoption of inappropriate levels of trust and subsequent behavioural adaptation. Nonetheless, when anticipating habituation it might also be necessary that important hard-warnings are introduced with a high urgency-level (for example our acoustic-tactile interface with illumination). Otherwise, habituation might weaken these signals over time to levels perceived as non-urgent. However, long-term investigations of these considerations are necessary to achieve more insight in this matter.

4.4.2. Behavioural adaptation

Our concepts provide continuous feedback on system-state and system changes. Drivers could easily perceive this information as confirmation of automation-activity and therewith degrade their barrier to perform secondary tasks and subsequently reduce their responsibility in supervision. Therefore, it is very important to bring across an appropriate level of trust in automation (Lee & See, 2004). Conveying information on system state is a strong advantage of the icon-concept (B), which displays the ego-vehicle and its surrounding situational features in a picture behind the steering wheel. Nonetheless, such advantages were not demonstrated during our short time-span tests. Further research is needed of how behavioural adaptation in the long-term is influenced by the particular features of the illumination-concept (C) as well as the icon-concept (B). Other important aspects that might influence behavioural adaptation are driver's attitude towards technology, willingness to take risks and driving experience.

4.4.3. Driving experience and training

Although the illumination-concept (C) helps to raise SA and showed superior performance in task-detection, none of the concepts improved intervention-support and the raised levels of SA are rather small. In other words, the offered support only has a small share in improving the renewed driving task, i.e. supervision with now and then intervention. Considering that usability is not an intrinsic product characteristic, but resulting from the combination of circumstances, product-interface and user-experience, it is advisable to review potential benefits of additional training to operate systems for automated driving. The potential benefits to offer intrinsic training, i.e. while using the concepts, are important areas for further research. Based on previous research, we expect that the mechanisms to raise performance from knowledge-based level to skill-based levels might offer valuable insights to comprise intrinsic training (Van den Beukel & Van der Voort, 2012).

5. Concluding remarks

This research tested three interface-concepts intended to support drivers with their changing role when driving partially automated (i.e. supervision with now-and-then intervention). Concept A provided only acoustic warnings and served as a baseline. Concept B presented status-information and take-over requests in a conventional manner: It displayed icons behind the steering wheel. Concept C was more advanced, using illumination in the windscreen and haptic feedback in the seat-pan to direct attention towards the locus of a potential hazard and to create an affordance for intervention when required. Concept C only combined illumination with vibrotactile feedback when intervention was needed. The illumination-concept (C) showed superior support for supervision: It showed significant better hazard-detection then the icon-concept (B) and hazard-detection is an important aspect for supervisory control. Moreover, measurement of Situation Awareness (another important aspect for supervisory control) denoted significant differences in favour of the illumination-concept (C) within two scenarios requiring supervision. Knowing that supervision will be the dominating driver's responsibility when utilising partially automated driving, we conclude that the illumination-concept (C) is a recommendable solution for support of the driver's changing role. Furthermore, it was considered that the illumination helps drivers to return more easily in the control-loop and to switch attention more easily between in-vehicle tasks and traffic-related circumstances. Nonetheless, neither the illumination-concept (C), nor the icon-concept (B) showed support for intervention. During intervention the illumination-concept (C) scored comparable or worse than the baseline-concept (A). Although these differences were not statistically significant, the scores are remarkable and against expectations. It was hypothesised that the combination of concept C's intervention-warning, especially the unexpectedness of the vibrotactile stimulus in addition to the acoustic and visual stimuli, caused counter-productive levels of annoyance and sometimes startle. With regard to overall support for the changing driver's role (including supervision and intervention) participants' subjective preferences are nonetheless in favour of the illumination-concept (C). This seems to confirm participants' recognition of the features that are intended to raise support. Furthermore, it is concluded that intervention and supervision benefit from different interface-features. Adding to this conclusion, we discussed possibly underlying causes for the ambiguity shown with concept C. For instance, how modality and the applied combination of stimuli might have affected annoyance or startle. The considerations provided in this paper contribute to further development of – and knowledge about – appropriate driver-vehicle interaction while vehicle-operation advances into operating partially automated driving systems, requiring supervisory-control and interventions.

Acknowledgements

This research was initiated and conducted at University of Twente. Co-financing of this research by Ford Motor Company is gratefully acknowledged.

References

Abbink, D. A., & Mulder, M. (2010). Motivation for a neuromuscular basis for haptic shared control. In *Paper presented at the analysis, design, and evaluation of human-machine systems*.

Ackerman, E. (2015). Tesla Model S. Summer software update will enable autonomous driving. *IEEE Spectrum Cars That Think*.

Anthony, B. J., & Putnam, L. E. (1985). Cardiac and blink reflex concomitants of attentional selectivity: A comparison of adults and young children. *Psychophysiology, 22*(5), 508–516.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775–779. http://dx.doi.org/10.1016/0005-1098(83)90046-8.

Baldwin, C. L., & Lewis, B. A. (2014). Perceived urgency mapping across modalities within a driving context. *Applied Ergonomics, 45*(5), 1270–1277. http://dx.doi.org/10.1016/j.apergo.2013.05.002.

Baldwin, C. L., & May, J. F. (2011). Loudness interacts with semantics in auditory warnings to impact rear-end collisions. *Transportation Research Part F-Traffic Psychology and Behaviour, 14*(1), 36–42. http://dx.doi.org/10.1016/j.trf.2010.09.004.

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine, 6*(4), 6–22.

Blumenthal, T. D. (1996). Inhibition of the human startle response is affected by both prepulse intensity and eliciting stimulus intensity. *Biological Psychology, 44*(2), 85–104.

Bohlin, G., Graham, F. K., Silverstein, L. D., & Hackley, S. A. (1981). Cardiac orienting and startle blink modification in novel and signal situations. *Psychophysiology, 18*(5), 603–611. http://dx.doi.org/10.1111/j.1469-8986.1981.tb01832.x.

Bouton, M. E. (2007). *Learning and behavior: A contemporary synthesis*. Sunderland, MA: Sinauer.

Braun, C. C., Sansing, L., Kennedy, R. S., & Silver, N. C. (1994a). Signal word and color specifications for product warnings: An isoperformance application. In *Paper presented at the proceedings of the human factors and ergonomics society annual meeting*.

Braun, C. C., Sansing, L., & Silver, N. C. (1994b). The interaction of signal word and color on warning labels: Differences in perceived hazard. In *Paper presented at the proceedings of the human factors and ergonomics society annual meeting*.

Brookhuis, K. A., De Waard, D., & Janssen, W. H. (2001). Behavioural impacts of advanced driver assistance systems–An overview. *European Journal of Transport and Infrastructure Research, 1*(3), 245–253.

Chapanis, A. (1994). Hazards associated with three signal words and four colours on warning signs. *Ergonomics, 37*(2), 265–275. http://dx.doi.org/10.1080/00140139408963644.

Charlton, S. G. (2002). Measurement of cognitive states in test and evaluation. In S. G. Charlton & T. G. O'Brien (Eds.), *Handbook of human factors testing and evaluation* (pp. 97–125). Mahwah, NJ: Lawrence Erlbaum Associates.

Daimler (Producer) (2015). *Active lane keeping assist for Mercedes-Benz passenger cars*. Retrieved from <www.daimler.com/Technology&Innovation>Safety>.

De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*(Part B), 196–217. http://dx.doi.org/10.1016/j.trf.2014.06.016.

Donges, E. (1999). A conceptual framework for active safety in road traffic. *Vehicle System Dynamics, 32*(2–3), 113–128. http://dx.doi.org/10.1076/vesd.32.2.113.2089.

Edworthy, J., Hellier, E., Walters, K., Clift-Mathews, W., & Crowther, M. (2003). Acoustic, semantic and phonetic influences in spoken warning signal words. *Applied Cognitive Psychology, 17*(8), 915–933. http://dx.doi.org/10.1002/acp.927.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors, 33*(2), 205–231. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-0025832576&partnerID=40&md5=afbac32cc2b2f1cd2530e8d22e4add49>.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic-systems. Human factors. *The Journal of the Human Factors and Ergonomics Society, 37*(1), 32–64. http://dx.doi.org/10.1518/001872095779049543.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462–492. http://dx.doi.org/10.1080/001401399185595.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. [See also (Kaber, Endsley, 1997): OOTL performance problems and the use of intermediate levels of automation for improved control system functioning and safety]. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(2), 381–394. Retrieved from <http://hfs.sagepub.com/content/37/2/381.short>.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the human factors and ergonomics society 42nd annual meeting* (Vols. 1 and 2, pp. 82–86). Retrieved from WOS:000079494900017.

EU (2011). *White paper: Roadmap to a single European transport area-Towards a competitive and resource efficient transport system*. European Commission Retrieved from <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011DC0144&from=EN>.

Fagerlonn, J. (2011). Urgent alarms in trucks: Effects on annoyance and subsequent driving performance. *IET Intelligent Transport Systems, 5*(4), 252–258. http://dx.doi.org/10.1049/iet-its.2010.0165.

Farrell, S., & Lewandowsky, S. (2000). A connectionist model of complacency and adaptive recovery under automation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*(2), 395.

Ferris, T., Penfold, R., Hameed, S., & Sarter, N. (2006). The implications of crossmodal links in attention for the design of multimodal interfaces: A driving simulation study. In *Paper presented at the proceedings of the human factors and ergonomics society annual meeting*.

Fitch, G. M., Hankey, J. M., Kleiner, B. M., & Dingus, T. A. (2011). Driver comprehension of multiple haptic seat alerts intended for use in an integrated collision avoidance system. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(4), 278–290. http://dx.doi.org/10.1016/j.trf.2011.02.001.

Flemisch, F. O., Bengler, K., Bubb, H., Winner, H., & Bruder, R. (2014). Towards cooperative guidance and control of highly automated vehicles: H-Mode and Conduct-by-Wire. *Ergonomics, 57*(3), 343–360. http://dx.doi.org/10.1080/00140139.2013.869355.

Gartenberg, D., McCurry, M., & Trafton, G. (2011). Situation awareness reacquisition in a supervisory control task. In *Paper presented at the proceedings of the human factors and ergonomics society annual meeting*.

Gasser, T. M., & Westhoff, D. (2012). BASt-study: Definitions of automation and legal issues in Germany. In *Paper presented at the presentation at the road vehicle automation workshop*.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? In *Paper presented at the proceedings of the human factors and ergonomics society annual meeting*.

Götze, M., & Bengler, K. (2015). Urban driving: Where to present what types of information–comparison of head-down and head-up displays human interface and the management of information. In *Information and Knowledge in Context* (pp. 190–200). Springer.

Green, P., Sullivan, J., Tsimhoni, O., Oberholtzer, J., Buonarosa, M. L., Devonshire, J., . . ., Sayer, J. (2008). *Integrated vehicle-based safety systems (IVBSS): Human factors and driver-vehicle interface (DVI) summary report*. Retrieved from.

Hellier, E., & Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. *Applied Ergonomics, 30*(2), 167–171. http://dx.doi.org/10.1016/S0003-6870(97)00013-6.

Ho, C., Tang, H. Z., & Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F-Traffic Psychology and Behaviour, 8*(6), 397–412. http://dx.doi.org/10.1016/j.trf.2005.05.002.

Hoc, J.-M., Young, M. S., & Blosseville, J.-M. (2009). Cooperation between drivers and automation: Implications for safety. *Theoretical Issues in Ergonomics Science, 10*(2), 135–160.

Hoffman, J., Lee, J. D., & Hayes, E. M. (2003). Driver preference of collision warning strategy and modality. In *Paper presented at the proceedings of the second international driving symposium on human factors in driver assessment, training and vehicle design*.

Infra, & Milieu. (2015). Chauffeur aan het stuur? *Ministerie van Infrastructuur en Milieu*. Retrieved from <http://www.kimnet.nl/sites/kimnet.nl/files/chauffer-aan-het-stuur.pdf>.

Jamson, A. H., Merat, N., Carsten, O. M. J., & Lai, F. C. H. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies, 30*, 116–125. http://dx.doi.org/10.1016/j.trc.2013.02.008.

Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress, 16*(3), 126–131.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113–153.

Kiefer, D., LeBlanc, D., Palmer, M., Salinger, J., Deering, R., & Shulman, M. (1999). *Development and validation of functional definitions and evaluation procedures for collision warning/avoidance systems*. Retrieved from <http://ntl.bts.gov/lib/jpodocs/repts_te/87l01!.pdf>.

Larsson, A. F. L., Kircher, K., & Hultgren, J. Andersson (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*(Part B), 229–237. http://dx.doi.org/10.1016/j.trf.2014.05.008.

Lee, J. D., Hoffman, J. D., & Hayes, E. (2004). *Collision warning design to mitigate driver distraction*.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(1), 50–80.

Lee, F. C. H., & Chan, A. H. S. (2007). Attending visual and auditory signals: Ergonomics recommendations with consideration of signal modality and spatial stimulus–response (S–R) compatibility. *International Journal of Industrial Ergonomics, 37*(3), 197–206.

Lee, H. K., Suh, K. S., & Benbasat, I. (2001). Effects of task-modality fit on user performance. *Decision Support Systems, 32*(1), 27–40. http://dx.doi.org/10.1016/S0167-9236(01)00098-7.

Lipp, O. V., Siddle, D. A., & Dall, P. J. (2000). The effect of warning stimulus modality on blink startle modification in reaction time tasks. *Psychophysiology, 37*(01), 55–64.

Liu, Y. C., & Jhuang, J. W. (2012). Effects of in-vehicle warning information displays with or without spatial compatibility on driving behaviors and response performance. *Applied Ergonomics, 43*(4), 679–686.

Marshall, D. C., Lee, J. D., & Austria, P. A. (2007). Alerts for in-vehicle information systems: Annoyance, urgency, and appropriateness. *Human Factors, 49*(1), 145–157. http://dx.doi.org/10.1518/001872007779598145.

Martens, M. H. (2004). Stimuli fixation and manual response as a function of expectancies. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(3), 410–423.

Martens, M. H. (2007). *The failure to act upon important information: Where do things go wrong*. Amsterdam, The Netherlands: Vrije Universiteit.

Martens, M. H., Pauwelussen, Schieben, Flemisch, F., Merat, Jamson, & Caci. (2008). *Human factors' aspects in automated and semi-automatic transport systems: State of the art*. Retrieved from.

Martens, M. H., & Van den Beukel, A. P. (2013). *The road to automated driving: Dual mode and human factors considerations. 16th international IEEE conference on paper presented at the intelligent transportation systems-(ITSC)*.

Merat, N., Jamson, A. H., Lai, F. C., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 54*(5), 762–771.

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 274–282. http://dx.doi.org/10.1016/j.trf.2014.09.005.

Merat, N., & Lee, J. D. (2012). Preface to the special section on human factors and automation in vehicles designing highly automated vehicles with the driver in mind. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 54*(5), 681–686.

Mulder, M., Abbink, D. A., & Boer, E. R. (2012). Sharing control with haptics seamless driver support from manual to automatic control. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 54*(5), 786–798.

Mulder, M., Van Paassen, M. M., & Abbink, D. A. (2008). Haptic gas pedal feedback. *Ergonomics, 51*(11), 1710–1720. http://dx.doi.org/10.1080/00140130802331583.

Naujoks, F., Purucker, C., Neukum, A., Wolter, S., & Steiger, R. (2015). Controllability of partially automated driving functions – Does it matter whether drivers are allowed to take their hands off the steering wheel? *Transportation Research Part F: Traffic Psychology and Behaviour, 35*, 185–198. http://dx.doi.org/10.1016/j.trf.2015.10.022.

Nielsen, J. (2005). *Ten usability heuristics*.

Nieto, M., Otaegui, O., Vélez, G., Ortega, J. D., & Cortés, A. (2015). On creating vision-based advanced driver assistance systems. *IET Intelligent Transport Systems, 9*(1), 59–66. http://dx.doi.org/10.1049/iet-its.2013.0167.

Nirschl, G. (2007). Human-centered development of advanced driver assistance systems human interface and the management of information. In *Interacting in Information Environments* (pp. 1088–1097).

Norman, D. A. (1990). The'problem'with automation: Inappropriate feedback and interaction, not'over-automation'. *Philosophical Transactions of the Royal Society B: Biological Sciences, 327*(1241), 585–593.

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*(Part B), 252–263. http://dx.doi.org/10.1016/j.trf.2014.04.009.

PReVENT. (2009). *Response 3 – code of practice for the design and evaluation of ADAS*. Retrieved from <http://www.acea.be/images/uploads/files/20090831 Code_of_Practice_ADAS.pdf>.

SAE_International. (2013). *SAE levels of driving automation*. <http://cyberlaw.stanford.edu/loda>.

Seppelt, B. D., & Lee, J. D. (2007). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies, 65*(3), 192–205. http://dx.doi.org/10.1016/j.ijhcs.2006.10.001.

Smisek, J., Van Paassen, M. M., Mulder, M., & Abbink, D. A. (2013). Neuromuscular analysis based tuning of haptic shared control assistance for UAV collision avoidance. In *Paper presented at the world haptics conference (WHC)*.

Stanton, N. A., Dunoyer, A., & Leatherland, A. (2011). Detection of new in-path targets by drivers using Stop&Go Adaptive Cruise Control. *Applied Ergonomics, 42*(4), 592–601. http://dx.doi.org/10.1016/j.apergo.2010.08.016.

Tan, A., & Lerner, N. D. (1995). *Multiple attribute evaluation of auditory warning signals for in-vehicle crash avoidance warning systems* Technical report. Springfield, VA: National Highway Traffic Safety Administration.

Toffetti, A., Wilschut, E. S., Martens, M. H., Schieben, A., Rambaldini, A., Merat, N., & Flemisch, F. (2009). CityMobil: Human factor issues regarding highly automated vehicles on eLane. *Transportation Research Record*, 1–9.

Van Arem, B., Van Driel, C. J. G., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems, 7*(4), 429–436. http://dx.doi.org/10.1109/tits.2006.884615.

Van den Beukel, A. P., & Van der Voort, M. C. (2016a). Drivers and automation – Does 'directional illumination' support mode awareness and understanding of driver's role? *IET Intelligent Transport Systems* (submitted for publication).

Van den Beukel, A. P., & Van der Voort, M. C. (2016b). How to assess driver's interaction with partially automated driving - A framework for early concept assessment. *Applied Ergonomics* (accepted for publication).

Van den Beukel, A. P., & Van der Voort, M. C. (2012). *Making automated driving support-method for driver-Vehicle task allocation*. Berlin: Human-centred Design for Intelligent Transport Systems (HUMANIST).

Van den Beukel, A. P., & Van der Voort, M. C. (2013). *The influence of time-criticality on situation awareness when retrieving human control after automated driving. 16th international IEEE conference on paper presented at the intelligent transportation systems-(ITSC), 2000–2005*.

Van den Beukel, A. P., & Van der Voort, M. C. (2016). Driving automation & changed driving task - Effect of driver-interfaces on intervention. In *Paper presented at the IEEE intelligent vehicles conference, Gothenburg*.

Van den Beukel, A. P., Van der Voort, M. C., & Eger, A. O. (2015). Towards a framework for testing drivers' interaction with partially automated driving. In *IEEE 18th international conference on paper presented at the intelligent transportation systems (ITSC)*.

Van der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies, 5*(1), 1–10. Retrieved from <http://www.hfes-europe.org/accept/accepttrc97OK.pdf>.

Van Driel, C. J. G. (2007). *Driver support in congestion - An assessment of user needs and impacts on driver and traffic flow* Doctoral thesis. The Netherlands: Universiteit Twente.

van Waterschoot, B. M., & van der Voort, M. C. (2009). Implementing human factors within the design process of advanced driver assistance systems (ADAS). In *Engineering psychology and cognitive ergonomics*. Springer.

Wiese, E. E., & Lee, J. D. (2004). Auditory alerts for in-vehicle information systems: The effects of temporal conflict and sound parameters on driver attitudes and performance. *Ergonomics, 47*(9), 965–986. http://dx.doi.org/10.1080/00140130410001686294.

Wogalter, M. S., Conzola, V. C., & Smith-Jackson, T. L. (2002). Research-based guidelines for warning design and evaluation. *Applied Ergonomics, 33*(3), 219–230. http://dx.doi.org/10.1016/S0003-6870(02)00009-1.

Wogalter, M. S., Kalsher, M. J., Frederick, L. J., & Magurno, A. B. (1998). Hazard level perceptions of warning components and configurations. *International Journal of Cognitive Ergonomics, 2*(1–2), 123–143.