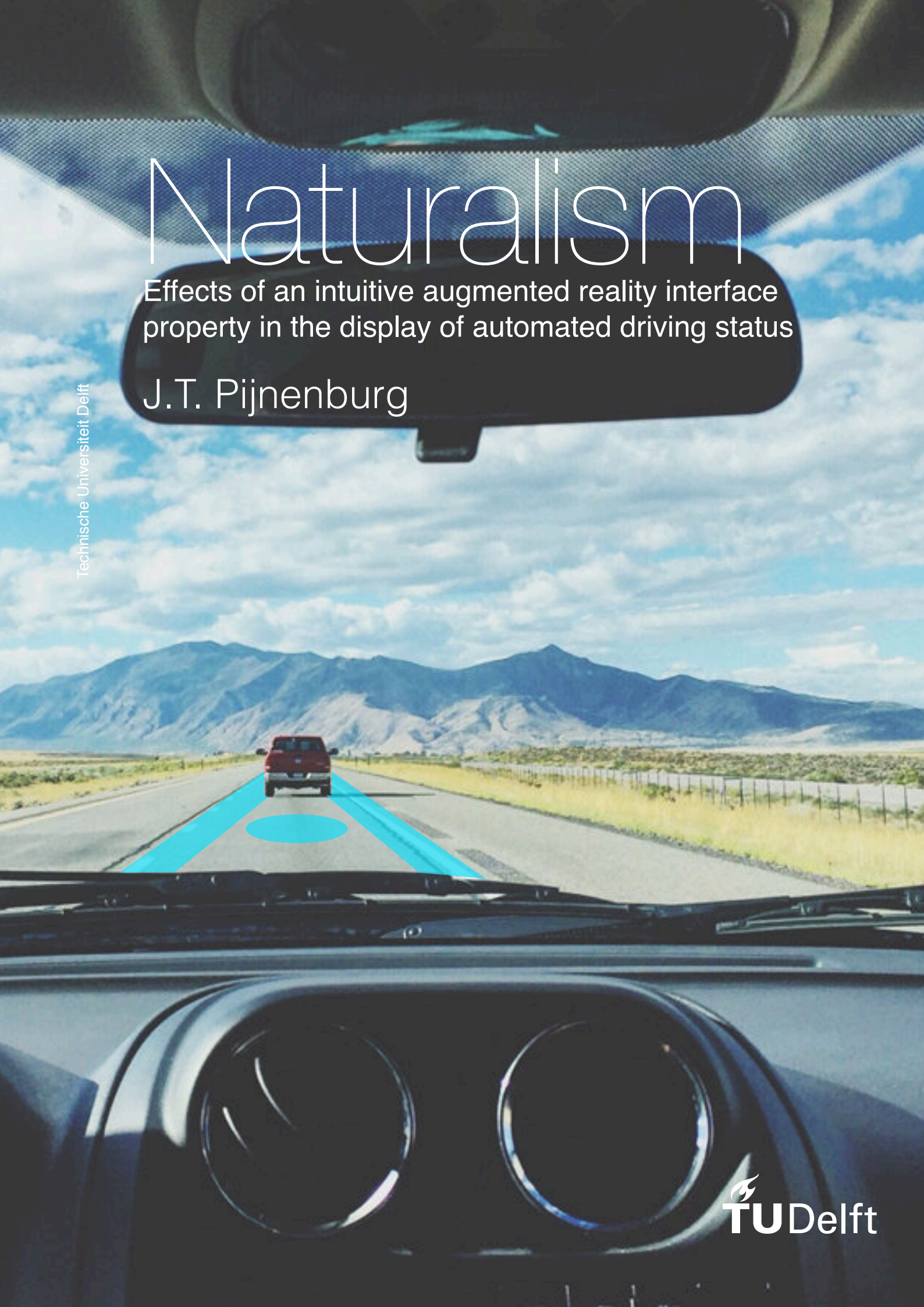


Naturalism

Effects of an intuitive augmented reality interface property in the display of automated driving status

J.T. Pijnenburg

Technische Universiteit Delft

TU Delft

Naturalism

Effects of an intuitive augmented reality interface property in the display of automated driving status

master thesis

By

J. T. Pijnenburg

4103157

in partial fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

at the Delft University of Technology,
faculty of Mechanical, Maritime and Materials Engineering

Chairman committee:	Dr. Ir. J.C.F. de Winter	TU Delft
Thesis committee:	Dr. Ir. R. Happee	TU Delft
	C.D.D. Cabrall, MSc	TU Delft
	Mgr. N. Kovácsová	TU Delft

Picture on the cover: Getty Images

TU Delft

Acknowledgement

Before you lies the master thesis "Naturalism: Effects of an intuitive augmented reality interface property in the display of automated driving status", which I defended December 15th, 2017.

I am grateful to C.D.D. Cabrall, who guided me with great passion, and to Dr. Ir. J.F.C. de Winter, for his expert advice. Thank you both for your dedication to this research and the opportunities of learning for me. Dr. Ir. R. Happee and Mgr. N. Kovácsová, I appreciate your effort as part of the Thesis committee.

I would like to specifically thank my mother Wendy, who I could always call for advice, and my sister Suzi, my father Jef, and also the rest of my family and friends. This accomplishment would not have been possible without them. Thank you.

Jonas

December 2017

Abstract

With various levels of automated vehicles on the verge, the human will be put in the operator role and this poses historic human-machine interaction challenges regarding sustained attention, mental workload, and engagement. Augmented reality based interfaces may help address some of these problems. We presume that more lifelike interfaces, in contrast to static icon based interfaces, could reduce the operators cognitive strain, by allowing him to store information externally and reducing cognitive switching between reality and interface. To evaluate augmented reality interfaces we introduce naturalism, as an interface property dimension ranging from arbitrary to natural

In this research, we question whether a more naturalistic interface could improve performance, vigilance, and subjective evaluation (workload, acceptance, and engagement). For this we tested an arbitrary, a semi-natural and a natural automation status interface.

These three interfaces were tested in a driving simulator setup using non-interactive real-life driving videos, to which 28 participants had to respond when an automation status error event occurred, which appeared after a precipitating cue such as appearing road construction.

Results seem to suggest that the semi-natural interface improves reaction time performance most, followed by the natural interface. With the natural interface vigilance seems to be improved, and mental demand is decreased, but all results should be interpreted with caution due to small sample sizes and other limitations. It seems that the dynamic response to precipitating cues, an intrinsic property of the semi-natural and natural interfaces, was of influence. The reduced mental demand seems to support our presumed mechanism of reduced cognitive strain. For further research, a more challenging task design and eye-tracking could be promising. To conclude, this research showed that more natural interfaces show potential to improve safety and comfort.

Contents

1 Introduction

Various levels of automated vehicles are on the verge[1 2]. As Parasuraman et. al. state, "It has become evident that automation does not supplant human activity; rather, it changes the nature of the work that humans do, often in ways unintended and unanticipanted by the designers of automation"[3]. Though driving functions of the vehicle are being automated, the human operator could still be responsible. These responsibilities occur in automated vehicles for SAE levels 2 and 3, for which the driving environment is monitored by the human operator and automated driving system respectively[4]. Even though the driving environment is monitored by an automated driving system for SAE level 3, the human operator should respond when the car crosses outside of the operational design domain, e.g. during roadworks. The role of the human in the vehicle changes. Bainbridge argues that automation and the human in an operator or supervisor role may increase difficulties[5]. Various SAE level 2 automated driving systems exist on-market, with different Human-Machine Interfaces (HMI) automation status displays.

1.1 Human error

When populations of vehicles grew, and the associated death and injury did also, research on Human Factors contributed to improve safety. Hancock states that "It is a tribute to the creativity of the technologists' art that fatality rates, and to an extent injury rates also, have not burgeoned at the proportionate rate that may have been anticipated."[6] Still, in 94 percent of all fatal crashes, the major factor is human error[7]. It seems that human error poses a serious challenge for interaction between man and machine. HMI challenges should be resolved before deploying automated vehicles on public roads[8]. A type of human error challenge, Vigilance (section 1.2), is explained in the next paragraph. In the next paragraphs we propose a solution in use of Visual HMIs (section 1.3) using Augmented reality (section 1.4), and Naturalism (section 1.5) as a specific property of these interfaces. In section 1.6, Presumed cognitive mechanism, explains how the three elements of the proposed solution could address human error and vigilance.

1.2 Vigilance

Vigilance can be defined as the ability to sustain attention to a task for a period of time in order to detect and respond to infrequent and non-salient events, and is also called sustained attention. Most theories on vigilance have focussed on decrement of performance. In 1948, Mackworth theorised that the decrease or absence of some form of positive stimulation is the cause of inhibition of performance[9]. Since Mackworth other early theories were formulated. The arousal theory suggests that people become disinterested and lose perceptual sensitivity due to the understimulating nature of the vigilance task. Arousal can be defined as the physiological and psychological state of being awake and reactive. Over time people become under-aroused, causing vigilance decrement. The arousal theory seems to explain part of the vigilance decrement, but is challenged by considering vigilance tasks as understimulating and mentally undemanding. As Warm et. al. point out, vigilance tasks are demanding on information-processing resources of subjects, and highly stressful[10]. In general, resource theories describe that mental resource processing assume for human performance in terms of supply and demand. This means tasks present costs on mental resources, and if those costs are relatively too high (for too long), performance suffers in consequence. Davies and Parasuraman noted that no singular theory would be able to account for vigilance decrement[11]. While vigilance theories are a nuanced matter, it seems that vigilance tasks pose human-machine interface challenges, as humans are put in the position to watch for an infrequent, temporally uncertain, unambigious, time-critical signal[12]. Hancock makes a strong statement about humans doing vigilance tasks: they are "magnificently disqualified"[6].

1.3 Visual HMIs

In this research we question wether visual HMIs can improve vigilance in monitoring automated driving. Visual human machine interface (HMI) displays of automation status are commonly employed by automobile manufacturers and could possibly improve vigilance and mode awareness. Beyond aesthetic choices in the automation status visualisation design space, functional benefits can also be included. For

example, a common functional design decision is to leverage spatial location and converge upon placing status information in line of sight of the driver, reducing the amount of time and effort spent looking away from the road[13]. An approach to reduce this time and effort is the head-up display (HUD), which shows augmented reality (AR) interfaces on the wind-screen. However, wind-screen located visualizations might pose potential dangers of information processing conflict. The design of information content and structure requires consideration. While on-market solutions use artificial information such as icons and text, automotive manufacturer future concepts and scholarly publications are presenting a wide array of interfaces using natural shapes and properties.

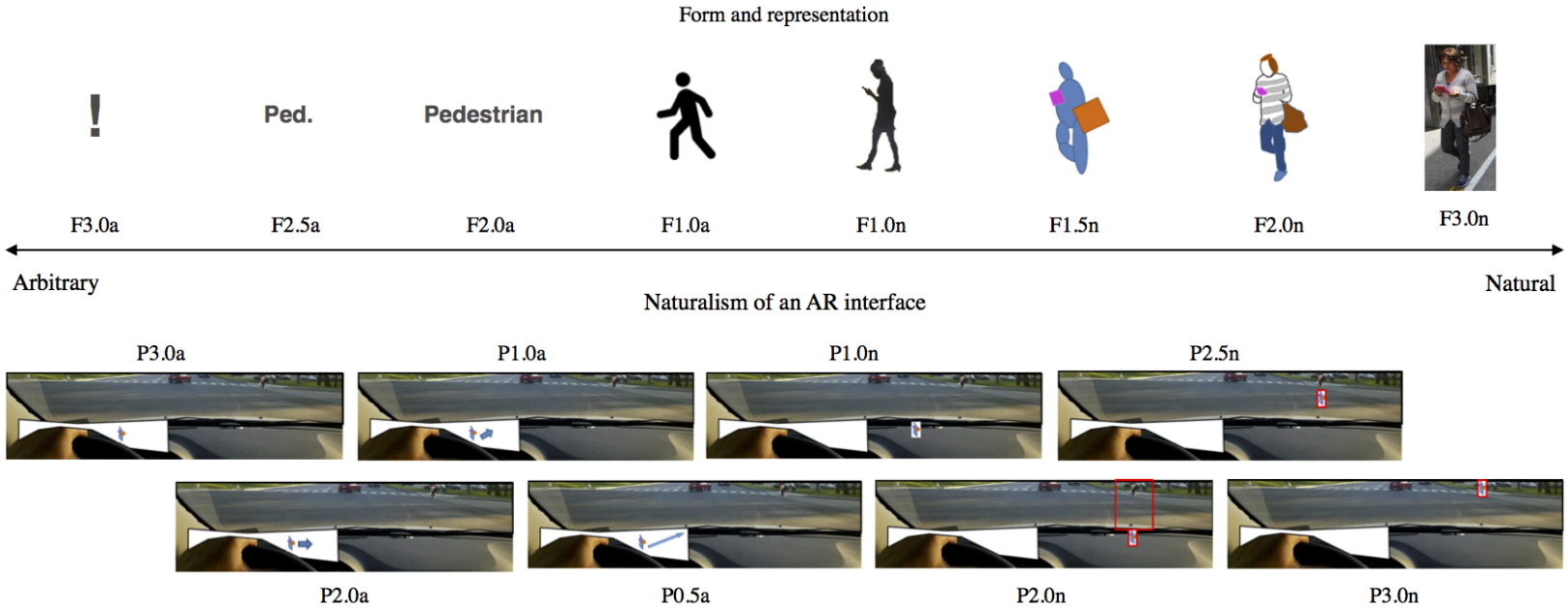
1.4 Augmented reality

Use of AR could show potential in aiding drivers in monitoring of automated driving. Currently automotive AR is in the research and development phase, which makes it a timely topic of interest. In their european roadmap Dokic et. al. predict AR to be in this phase until 2020, then transitioning to a demo phase and hitting product and market from 2022[14]. Augmented reality is defined as a variation of virtual reality, in which the user can see the real world with virtual objects superimposed or composited with the real world[15][16]. AR can highlight both hidden affordances and perceptible affordances[17]. The actor receives focused information enabling him to act upon this affordance. The goal of this study is to research how different AR based interfaces and interface properties impact performance, workload and acceptance during a vigilance-type task in a highly automated vehicle. This could lead to safer and more enjoyable human-automobile interaction in the future.

1.5 Naturalism

In contrast to the dashboard, which conventionally has informed the user by showing arbitrary icons, AR can emphasize reality. To be able to discuss and investigate potential advantages of highlighting actual real life cues in a vivid, natural way, we introduce a new scale to describe this quality of an AR interface. This scale describes differences in interfaces that leverage real-life cues, using real-life motion, shapes, depth-perception and intuitive stimuli versus more arbitrary interfaces with components that rely on more symbolic information such as knowledge-based icons, numbers and gauges, etc.
Our literature review on automotive augmented reality includes an overview of AR interfaces in recent literature, shows how they rank on the Arbitrary to Natural scale, and argues that current on-market HUDs are can be classified as Arbitrary AR interfaces, and can be found in Appendix A. Figure 1 shows examples from recent literature arranged on this scale.
To further specify this scale, we distinguish form and representation from position and dynamics. In Figure 1, a blend is shown. Figure 2 further illustrates these sub qualities of the arbitrary-natural dimension, using a interface indicating a pedestrian as an example. Note that this is an example, other divisions, levels and intervals exist. This Figure also depicts that a more natural interface intrinsically contains more information.

Figure 1: Examples of AR HMIs from recent literature along arbitrary-natural dimension



Form and representation		Position and dynamics		
F3.0a	Symbol	P3.0a	Visual is unaffected by pedestrian's actual position and movement	e.g. non-directional indication
F2.5a	Abbreviated text	P2.0a	Visual is minorly affected by pedestrian's actual position and movement	e.g. 1 dimensional indication
F2.0a	Text	P1.0a	Visual is somewhat affected by pedestrian's actual position and movement	e.g. multi dimensional indication, arrow indication depicts no depth clues
F1.0a	Icon	P0.5a	Visual is somewhat affected by pedestrian's actual position and movement	e.g. multi dimensional indication, arrow depicts depth clues
F1.0n	Flat textures, colors and shapes, corresponding to the specific pedestrian	P1.0n	Visual is affected by pedestrian's actual position and movement	e.g. located and moving more closely with pedestrian
F1.5n	Generic depth cued through rough textures, colors and shapes, corresponding to the specific pedestrian	P2.0n	Visual is loosely coupled to pedestrian's position and movement	e.g. nearby, surrounding the pedestrian
F2.0n	Specific depth cued through detailed textures, colors and shapes, corresponding to that specific pedestrian	P2.5n	Visual is closely coupled to pedestrian's position and movement	e.g. below the pedestrian
F3.0n	Full image of that specific pedestrian	P3.0n	Visual is directly coupled to pedestrian's position and movement	e.g. direct overlay of pedestrian position and movement

Figure 2: Naturalism scale for Form and Representation (top) and Position and dynamics (bottom)

The term Naturalism was chosen as the preferred terminology. While not perfect it seems sufficiently fitting for now as it is somewhat self-explanatory and emphasizes the contrast between arbitrary and natural properties. The main criteria were self-explanatory value and potential for confusion with existing concepts. The term 'level of augmentation', also has self-explanatory merits but is too easily misunderstood as this term could also relate to Milgram's Virtuality Continuum, which runs from real environment to virtual environment, with augmented reality inbetween[18]. The term 'skeuomorphism' was also considered, but the definition related too much to using old properties of previous design solutions. Old design properties help intuitive use in a cultural, knowledge-based way, which doesn't contrast with static arbitrary icons the way a natural intuitive design does, thus this term was found insufficiently fitting.

1.6 Presumed cognitive mechanism

Theories from scientific (human factors) research converge upon insights towards improved human machine interaction through more naturalistic interfaces. Specifically, the related work on ecological perception, situated approaches, and extended cognition present opportunities to enhance human monitoring of automated driving control.
Chiappe et. al. argue situation awareness (SA) to be stored both internally (your mind) and externally (the world). A natural interface could potentially emphasize affordances which helps the driver to offload situation awareness information externally. Chiappe et. al. state "The operator internally represents where to access it from the world and can do so quickly"[19]. This perspective is in line with Clark et. al, who suggest that the linking of a human with an external entity creates a cognitive system, which they call active externalism[20]. This helps alleviate cognitive processes externally. Alleviated cognitive processes could mean a better reaction time (RT) performance baseline and less vigilance decrement as less effort and more predictive support would be provided to maintain or renew SA. The cognitive effort from switching between the interface and reality is more entailed with abstract interfaces and less with natural interfaces. Presumably switching from a natural interface to reality has a lower cost of attention, thus drivers are more likely to pick up on important driving scene cues earlier, by keeping people's mind more directly within the driving scene.
In essence, the larger theoretical postulate is that a more natural interface, by being more intuitive, helps improve situation awareness and alleviate cognitive effort thus improving RT and accuracy, and decreasing workload.

1.7 Research aims and questions

The aim of this study was to simulate SAE level 2 and 3 status error detection tasks in an autonomous vehicle, using AR interfaces of automation status with different forms of naturalism to rougly reflect low, medium and high levels (i.e., arbitrary, semi-natural, natural). Two Experiment sessions called Exp. 1 (SAE level 3) and Exp. 2 (SAE level 2) will be further elaborated in the Methods section. The impact of the various interfaces was assessed through objective and subjective evaluation. An overview of three testable and theoretically relevant hypotheses can be seen in Table 1, regarding benefits of naturalistic AR toward increased reaction and response performance, decreased vigilance decrement effects and enhanced subjective user experience. These are split into 11 sub-hypothesis and further elaborated below.

The first sub-hypotheses are concerned with reaction performance, for both SAE level 3 and SAE level 2. If H1 proves to be true, it suggests that more naturalistic AR automation status displays may have potential to improve (baseline) performance.

- **H1-1/2-a:** *Humans respond more quickly using a more naturalistic augmented reality interface, when monitoring automation status, for both autonomous vehicle SAE level 2 and 3.*
- **H1-1/2-b:** *Humans respond more accurately using a more naturalistic augmented reality interface, when monitoring automation status, for both autonomous vehicle SAE level 2 and 3.*

The next sub-hypotheses are concerned with vigilance decrement, which we only test for SAE level 2, where prolonged and sustained attention is required of the operator per the defined SAE levels. If H2

proves to be true, this suggests that a more naturalistic AR automation status display may have potential to mollify the classically problematic vigilance decrement.

- **H2-2-a/b-i** *Human vigilance performance decrement in terms of increased reaction time and decreased hitrate is postponed* (offset in time) *using a more naturalistic augmented reality interface, when monitoring automation status, for autonomous vehicle SAE level 2.*
- **H2-2-a/b-ii** *Human vigilance performance decrement in terms of increased reaction time and decreased hitrate reduced* (offset in size) *using a more naturalistic augmented reality interface, when monitoring automation status, for autonomous vehicle SAE level 2.*

Lastly, we consider subjective evaluation of acceptance, workload and engagement, which are also tested for SAE level 2.

- **H3-2-i:** *Self-reported workload as measured by the NASA-Task Load Index (TLX) is lower using a more naturalistic augmented reality interface, when monitoring automation status, for autonomous vehicle SAE level 2.*
- **H3-2-ii:** *Acceptance as measured by the Van Der Laan acceptance questionnaire is higher using a more naturalisti c augmented reality interface, when monitoring automation status, for autonomous vehicle SAE level 2.*
- **H3-2-iii:** *Engagement as measured by the Dundee Stress State Questionnaire (DSSQ) short engagement section is higher using a more naturalistic augmented reality interface, when monitoring automation status, for autonomous vehicle SAE level 2.*

	Hypothesis category	Source	ID	Description
H1	Performance	Exp. 1	H1-1-a	Decrease reaction time
		Exp. 1	H1-1-b	Decrease response error
		Exp. 2	H1-2-a	Decrease reaction time
		Exp. 2	H1-2-b	Decrease response error
H2	Vigilance decrement	Exp. 2	H2-2-a-i	Delay onset of decrement over time of reaction time
		Exp. 2	H2-2-b-i	Delay onset of decrement over time of accuracy
		Exp. 2	H2-2-a-ii	Reduce decrement over time of reaction time
		Exp. 2	H2-2-b-ii	Reduce decrement over time of accuracy
H3	Self report	Exp. 2	H3-2-i	Improved workload
		Exp. 2	H3-2-ii	Improved acceptance
		Exp. 2	H3-2-iii	Improved engagement

Table 1: Hypothesis 1-3

2 Method

2.1 Participants

Student participants were provided through the course ME41000 at Delft University of Technology. 28 people participated in Exp. 1, the first part of the study. From the Exp. 1 participant pool, 18 participants were also randomly selected for participation in Exp. 2. Participants ages of Exp. 1 ranged between 20 and 28 years, with an average of 23.8. 24 were male, 4 were female (86.2 percent versus 13.8 percent). All but two had obtained a drivers licence, on average at age 18. Most participants drove 1 to 3 times per week, or once a month to once a week. Participants drove on average 6034 km yearly. Participants ages of Exp. 2 ranged between 20 and 28, with an average of 24.1. 15 were male, 3 were female (83.3 percent versus 16.7 percent). All but two had obtained a drivers licence, on average at age 18. Most participants drove once a month to once a week. Participants drove on average 5278 km yearly.

2.2 Apparatus

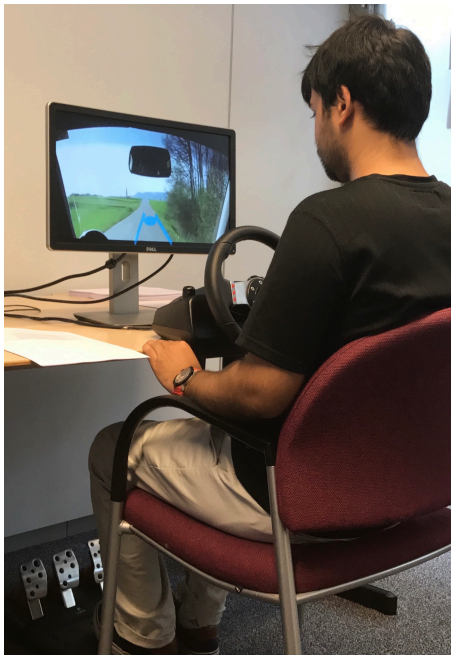


Figure 3: Non-interactive driving simulation setup with real-world driving videos

The experiment apparatus consisted of a gaming steering wheel and a stimulus display monitor (Figure 3). The gaming steering wheel was the model G27 by Logitech, with a wheel diameter of 280mm, and a range of rotation up to 900 degrees. The wheel featured 2 winged paddleshifters and 6 red buttons (which were used to navigate on-screen instructions). The foot pedals include an accelerator with a light spring, a brake with a heavy spring, and a clutch with a medium spring. The stimulus display monitor was a Dell P2414H with a display measuring 24 inch. It was an IPS LED wide-screen monitor. Paper questionnaires were used.

2.3 Stimuli

2.3.1 Driving videos

The study was conducted with pre-recorded real life driving videos, collected from Youtube. Driving video based data sets are not widely used in published automotive HMI studies, but were appealing because augmenting lifelike, realistic aspects of driving is key for naturalism. These visual stimuli were required to be of similar and relatively high visual quality. Videos with appropriate camera position were selected, appropriate being near first-person point of view, fitting the full windscreen. Weather and driving environment were kept fairly constant, as all videos were sourced from the same Youtube contributor (e.g., same camera, mounting position, vehicle, local road, traffic, and geographic climate conditions). From 1 hour, 28 minutes and 38 seconds of source video 57 different videos were cut that included precipitating cues to increase error event likelihood. Audio was removed. A portion of these videos were purposefully selected to include precipitating cues that would conceivably increase likelihood of automation error event (i.e., status drop out) within the following categories: sharp turns, uneven roads, close objects, and road construction. Such categories were consistent with functional limitations as described in the on-screen instructions provided to the participants before the start of the first session. These videos include precipitating clues that offer additional information being highlighted by the interface, allowing the participant to externally offload situation awareness and form a cognitive system.
Adobe After Effects CC 2017 video editing software

was used to create and overlay visual interfaces (and interface property variations) on top of the driving video footage. These interfaces showed automation status of lateral and longitudinal position of the vehicle, which was either on or off.

2.3.2 Automation status error events

Participants were tasked with reacting as quickly and accurately as possible to automation error events (i.e., lateral and/or longitudinal status information disappearances). We defined these error events as a single point in time when the automation "on" status (lateral or longitudinal) disappeared, several seconds after a precipitating cue (Figure 5) being visible. The automation status error events were presented silently without any auditory cues.
Responses were made using a Logitech G27 gaming steering wheel and pedal set (Figure 3). The instructed response sequence was to first grab wing trigger handles behind the wheel (used to test RT), before then registering some press on the steering wheel or pedals. The intent was to ensure for participants to respond with realistic care and thought, and not mindlessly make responses, falsely and without urgency.

2.3.3 Interface conditions

Three interfaces were designed, as shown in Figure 4. The stripes on left and right corresponded to lateral automation status, i.e., detection of lane boundary edge lines. For the Arbitrary interface these were static, for the Semi-Natural interface these bent corresponding to the curve of the road, and for the Natural interface, these lines dynamically overlaid the lane boundary lines in a 3-D depth compatible manner.
The dot in the middle corresponded to longitudinal automation status, which indicated whether the vehicle was able to detect objects in front, and the distance to this object or lead vehicle, i.e., detection of objects and distances in front. In the Arbitrary interface this dot was static. In the Semi-Natural interface the dot moved up or down depending on the distance of an object in front, with upwards movements corresponding to increasing object distances and downwards movements corresponding to decreasing object distances. For the Natural interface an elliptical disk was used to convey natural depth perspective, which moved towards the object in front or towards the viewer, with movements towards the object corresponding to increasing object distances and movements towards the viewer corresponding to decreasing object distances.
With the intent of isolating effects of HMI arbitrariness and naturalism, attempts to control saliency were included in the display stimulus design. To control for a potential confound of stimulus saliency, the color, opacity, and approximate position and amount of HMI element pixels were kept as similar as possible although varied somewhat due to the previously described differences relatively intrinsic to the target concept of HMI arbitrariness and naturalism. Examples of the HMI interface in various driving situations can be found in Figure 6.

For Exp. 1, the conditions, videos and error events are randomised using a script in Matlab. For Exp. 2, conditions (ABC) and video scenario (123)

Condition A: Arbitrary Condition B: Semi Arbitrary-Natural Condition C: Natural

Figure 4: Interface designs ranging from arbitrary to natural

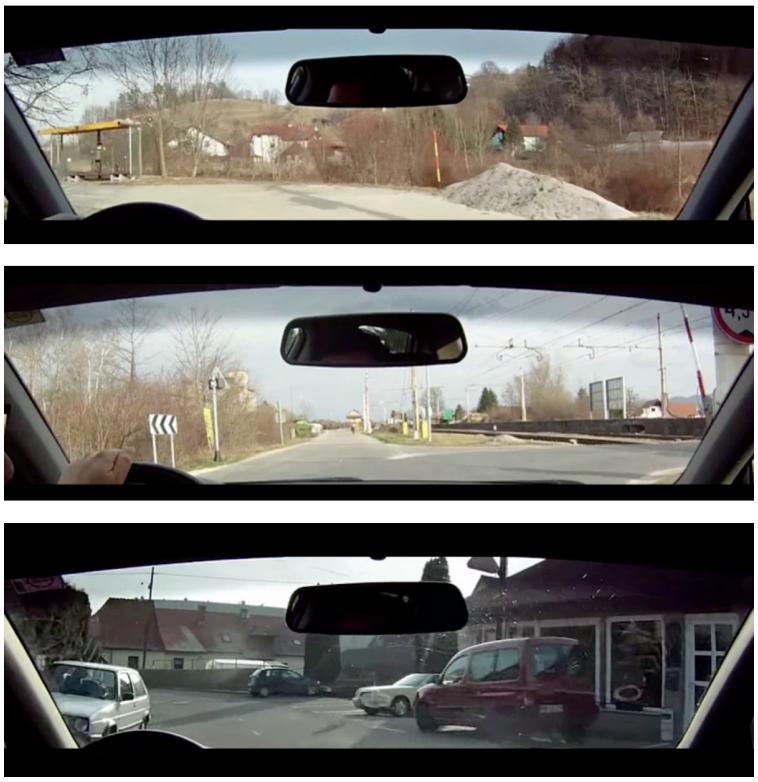


Figure 5: Three examples of precipitating cues for error events, from Exp 1. The top image shows the cue "Uneven roads", as it is a dirt road with construction material on the side. The middle image shows the cue "Sharp turn", at the railway crossing a very sharp turn has to be made. The bottom image shows the cue "Close object", as the red parked car appears very close.

were balanced to avoid learning effects and to test the interface, not the specific video. These combinations were balanced across slot tests (Exp2(1), Exp2(2), Exp2(3)) to deter temporal order bias effect. Using these criteria 9 combinations were selected, as shown in Table 2. Note that this is not fully balanced against order bias effects.

Combination	Exp2(1)	Exp2(2)	Exp2(3)
1	A1	B2	C3
2	C1	A2	B3
3	B1	C2	A3
4	A3	B1	C2
5	C3	A1	B2
6	B3	C1	A2
7	A2	B3	C1
8	C2	A3	B1
9	B2	C3	A1

Table 2: Overview of 9 combinations, of Interface Conditions [A-C] and Videos[1-3], for the order of Exp. 2

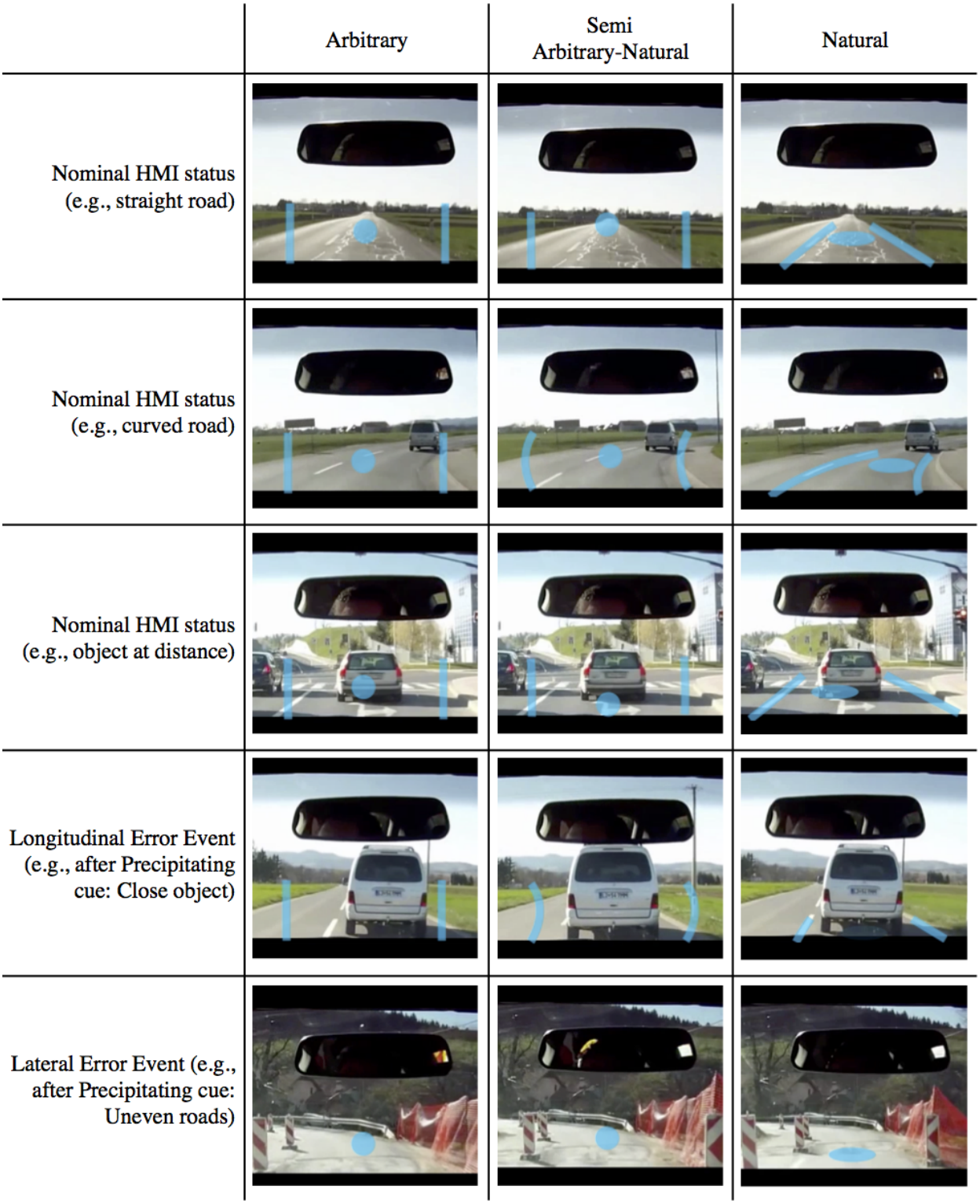


Figure 6: Arbitrary, Semi-Natural, and Natural interfaces across various example driving situations

	Number of participants	Design	Number of repetitions per condition	Duration	Simulates
Exp. 1	28	Within subjects	2	54 x 5-second videos	SAE level 3
Exp. 2	18	Within subjects	5 or 6	3 x 5-minute videos	SAE level 2

Table 3: Overview of Exp. 1 and 2. Exp. 1 employs short videos to simulate entering the active loop from a non-monitoring task (SAE level 3), while Exp. 2 employs longer videos in which the driver does actively monitor the environment (SAE level 2)

2.4 Exposure experiment sessions

Two experiments of exposure sessions were conducted, one representative of SAE level 3 and one representative of SAE level 2. These two exposure experiments correspond to the Exp. 1 and Exp. 2 in Table 3, 4 and 5.

For Exp. 1, 54 videos of 5 seconds were selected, accounting for 5 minutes and 24 seconds including 1-second black screens in-between videos. 17 Error events occurred in Exp. 1, with three different precipitating clues: Uneven roads (6), Sharp turn (5) and Close object (6). Examples of these three types of precipating cues are shown in Figure 5. For Exp. 2, three 5-minute videos were selected. Six error events occurred in Exp. 2, with three different precipating clues: Uneven roads (1), Sharp turn (2) and Close object (3).

A pilot study was conducted to validate use of driving videos and provide proof of concept the use of video-edited interfaces. Both were found to be realistic by all 6 participants (none of whom served as Exp. 1 or Exp. 2 participants). Video duration was determined through trial and error during the pilot test. Video durations (as validated during pilot testing) are provided in Table 3 and described below.

1. **Exp. 1.** 17 automation status error events were randomly ordered within 54 five-second duration videos (i.e., 0 or 1 error event per video) over the course of a 5 minutes and 23 seconds exposure session. Exp. 1 followed a fixed randomized order, which was pre-determined and equal for all participants. 1-second black screens were shown as inter-stimulus intervals. Participants were tasked to attend to a driving situation to which they had previously not be attending, as a new video was presented every 5 seconds. This was chosen to reflect a condition similar to that as might be commonly expected with SAE level 3 driving automation for which drivers are permitted to be out of and return to the loop of monitoring driving progress.
2. **Exp. 2.** 6 automation status error events were semi-randomly ordered within 3 videos of 5 minute duration (i.e., 2 error events per video) over the course of a 21 minute session with about 2 minutes break in between Exp 2. videos to fill in subjective questionnaires. Participants were required to monitor the environment and automated driving status full-time (compatible to SAE level 2) and vigilance decrement may likely occur.

2.5 Procedure

First, informed consent was provided and obtained via a written form (Appendix B). Participants sat in front of the display and were given instructions on-screen (Appendix C). Exp. 1 was conducted. A randomized subset of participants continued to Exp. 2, for which instructions appeared on-screen. Exp. 2 then was conducted, starting with Exp. 2(1), with questionnaires following. This was repeated for Exp. 2(2) and Exp. 2(3). The participants were thanked and received an oral debrief in which the purpose of the study was further explained and additional questions could be asked. Table 6 shows an overview with approximate timing.

2.6 Dependent measures

For Exp. 1, Accuracy and RT were measured. Reaction time (RT) is defined as the time between the error event and the stimulus response, for which the participant presses one or both winged paddleshifters (Figure 7). In regards to accuracy, responses were categorised into Hits, Misses and False Alarms. A

Video	Error type	Start(s)	End(s)	Error event(s)	Condition	Precipitating cue
1		0	5		Natural	
2		6	11		Semi-natural	
3		12	17		Natural	
4		18	23		Semi-natural	
5		24	29		Arbitrary	
6	Lateral	30	35	34.060	Arbitrary	Sharp turn
7		36	41		Natural	
8	Longitudinal	42	47	43.003	Arbitrary	Sharp turn
9	Lateral	48	53	49.068	Natural	Uneven roads
10	Lateral	54	59	57.002	Natural	Sharp turn
11		60	65		Semi-natural	
12		66	71		Semi-natural	
13		72	77		Semi-natural	
14		78	83		Arbitrary	
15	Longitudinal	84	89	87.002	Arbitrary	Close object
16	Lateral	90	95	92.002	Semi-natural	Sharp turn
17		96	101		Natural	
18		102	107		Arbitrary	
19		108	113		Natural	
20		114	119		Semi-natural	
21	Longitudinal	120	125	124.002	Semi-natural	Close object
22		126	131		Arbitrary	
23		132	137		Natural	
24		138	143		Arbitrary	
25		144	149		Arbitrary	
26	Lateral	150	155	151.050	Natural	Uneven roads
27		156	161		Natural	
28		162	167		Natural	
29		168	173		Natural	
30	Lateral	174	179	176.002	Arbitrary	Sharp turn
31		180	185		Arbitrary	
32		186	191		Arbitrary	
33	Lateral	192	197	194.002	Semi-natural	Uneven roads
34		198	203		Natural	
35	Longitudinal	204	209	204.052	Arbitrary	Close object
36		210	215		Semi-natural	
37	Lateral	216	221	217.052	Semi-natural	Uneven roads
38	Longitudinal	222	227	225.002	Semi-natural	Close object
39		228	233		Semi-natural	
40		234	239		Semi-natural	
41		240	245		Semi-natural	
42		246	251		Arbitrary	
43		252	257		Natural	
44	Longitudinal	258	263	261.002	Natural	Uneven roads
45	Longitudinal	264	269	268.052	Natural	Close object
46		270	275		Arbitrary	
47		276	281		Semi-natural	
48		282	287		Natural	
49	Longitudinal	288	293	291.002	Semi-natural	Close object
50		294	299		Arbitrary	
51		300	305		Natural	
52	Longitudinal	306	311	308.002	Arbitrary	Uneven roads
53		312	317		Semi-natural	
54		318	323		Arbitrary	

Table 4: Overview of Exp. 1. Note that 54 Videos were stitched together with 1-second black screen interval between each segment. Timings are provided above relative to the start of the full video.

Video	Error type	Start(s)	End(s)	Error event(s)	Condition	Precipitating cue
1	Longitudinal	0	300	223.002		Close object
1	Lateral	0	300	235.085		Sharp turn
2	Longitudinal	0	300	69.002		Road construction
2	Lateral	0	300	275.002		Close object
3	Longitudinal	0	300	158.002		Sharp turn
3	Lateral	0	300	239.052		Close object

Table 5: Order of driving scene video segments alternated between three different clips as a between subject grouping variable.

Exp. 1 only (n = 28)	**Approx. Duration**
Consent form	2m
Instructions on screen: 1	5m
Exp. 1: {54 videos, 17 error events}	5m
Debrief	1m
Total time	13m

Exp. 1 followed by Exp. 2 (n = 18)	**Approx. Duration**
Consent form	2m
Instructions on screen: 1	5m
Self-report set 1: {DSSQ}	1m
Exp. 1: {54 videos, 17 error events}	5m
Instructions on screen: 2	1m
Exp. 2(1): {1 video, 2 error events}	5m
Self report set 2: {DSSQ, Van der Laan, NASA-TLX}	2m
Exp. 2(2): {1 video, 2 error events}	5m
Self report set 3: {DSSQ, Van der Laan, NASA-TLX}	2m
Exp. 2(3): {1 video, 2 error events}	5m
Self report set 3: {DSSQ, Van der Laan, NASA-TLX}	2m
Debrief	1m
Total time	36m

Table 6: Procedural sequence for participants of both Exp. 1 and Exp. 2 exposure sessions of present research.

response was categorised as a Hit if the participant responded by pressing one or both winged paddleshifters within 3.5 seconds after the error event occurred. If the winged bars were pressed multiple times within these three seconds this was interpreted as a single response. If no response was given within 3.5 seconds of the error event, this was categorised as a Miss. Responses that were given beyond the 3.5 second window after an error event, were categorised as False Alarms. In regards to RT, responses were measured in milliseconds, from the time of the error event occurrence until the winged paddleshifters were pressed.

For Exp. 2, Accuracy and RT were measured and the same categorisation and measurement apply. In addition, workload, acceptance and engagement were measured. This was done using the NASA-TLX questionnaire, the Acceptance questionnaire[21] and the engagement components of the DSSQ short[22] (Appendix D). These questionnaires are answered three times. After the first video, after the second video and after the third video. The DSSQ short also has a baseline questionnaire which is asked before Exp. 1. No subjective measures were taken during Exp. 1 as the benefit of a compact uninterrupted, rapid succession of videos outweighed the value of an additional analysis when we expected a comparable representation in a more full task fidelity experience of simulated SAE 2.

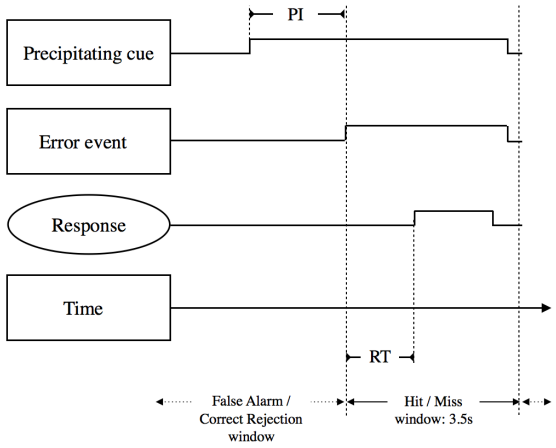


Figure 7: The time lines for error events in the typical procedure for measuring RT. PI, the Precipitating interval, varies per error event. RT, reaction time.

3 Results

3.1 H1 Performance

3.1.1 Exp. 1

Over the course of Exp. 1, each of the 28 participants were shown 17 error events for a total of 476. Of those 476 error events, 168 were with the Arbitrary interface, 168 were with the Semi-Natural interface, and 140 were with the Natural interface. 4 responses were removed because the participant misunderstood the exercise.

H1-1-a: Reaction time

For the average RT of 447 hits from 28 participants, the Semi-natural ($M = 689$ ms, $SD = 214$ ms) condition performed quicker than the Arbitrary ($M = 876$ ms, $SD = 188$ ms) and Natural ($M = 931$ ms, $SD = 179$ ms) interface conditions, as shown in Figure 8. Average RTs between error events differ, as shown in Figure 9. The average RT for error event 11, 1416 ms, is over a factor 3 compared of the average RT for error event 16, 453 ms. Due to the lack of balancing and difference in RT between error events, no further statistical tests were conducted.

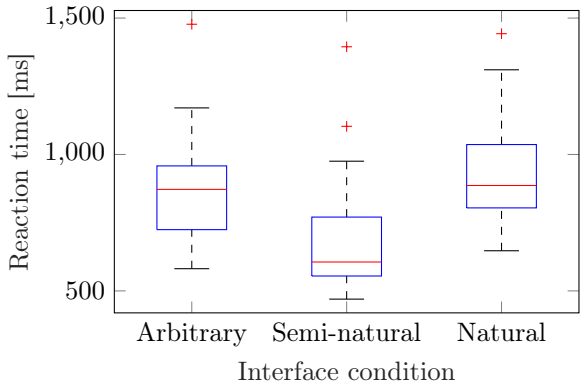


Figure 8: Reaction time for three conditions, Exp. 1

H1-1-b: Accuracy

Over the course of Exp 1, across 28 participants, the winged paddleshifters were pressed 463 times, of which 16 were false alarms and 447 were hits. 25 misses occurred. The participants' mean (SD) hit rate was 93.5%(11.4%), 96.43%(7.0%) and 94.3%(13.2%), whereas the mean (SD) false alarm rate was 2.0%(5.1%), 4.0%(7.2%) and 2.4%(5.9%) for the Arbitrary, Semi-Natural and Natural conditions respectively. No further statistical analysis was conducted for the reason explained in H1-1-a.

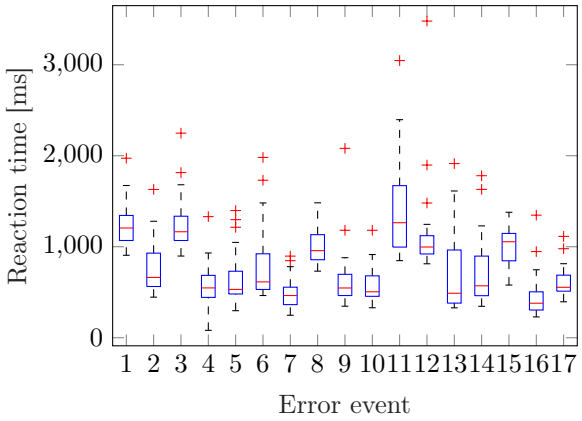


Figure 9: Reaction time for error event 1-17, Exp. 1

3.1.2 Exp. 2

Over the course of Exp. 2, each of the 18 participants were shown 6 error events for a total of 108. Of those 108 error events, 36 were with the Arbitrary interface, 36 were with the Semi-Natural interface, and 36 were with the Natural interface.

H1-2-a: Reaction time

Over the average RT of 106 hits (the other 2 being misses) it was observed that the Semi-natural ($M = 786$ ms, $SD = 397$ ms) condition showed a shorter average RT than the Arbitrary ($M = 1084$ ms, $SD = 424$ ms) and Natural ($M = 847$ ms, $SD = 270$ ms) conditions, as shown in Figure 11.

The mean RTs were found to be unequal according to a repeated measure one-way ANOVA, $F(2,34) = 5.28$, $p = 0.01$. The ANOVA was followed by paired comparisons between the three conditions using a Bonferroni correction. Hence, in the pairwise comparisons, a result was declared statistically significant if the p value was smaller than 0.05/3. The pairwise comparisons indicated no significant comparison. Arbitrary vs. Semi-natural: $p = .07$, Arbitrary vs. Natural: $p = .08$, Semi-natural vs. Natural:

$p = 1.0$. The difference between the pairwise comparisons p values suggest that the unequality as found by one-way repeated measures ANOVA is most likely due to the Arbitrary vs. Semi-natural and Arbitrary vs. Natural comparisons. More in-depth results are shown in Figure 10, 12 and 13.

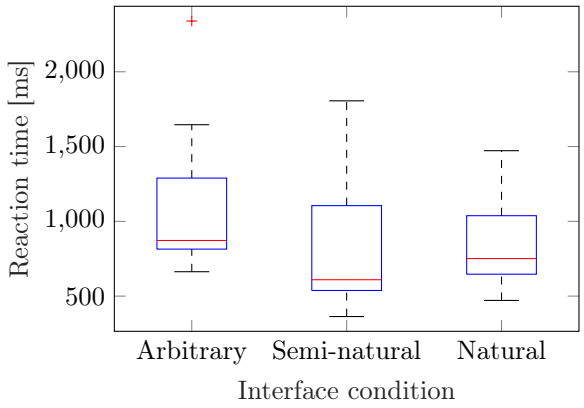


Figure 11: Reaction time for three conditions, from Exp 2. repeated measure one-way ANOVA, $F(2,34) = 5.28$, $p = 0.01$

H1-2-b: Accuracy

During Exp 2, the 18 participants pressed the winged paddleshifters 112 times, of which 6 were false alarm and 106 were hits. 2 misses occurred. The two misses occurred for the Arbitrary (1) and Natural (1) condition. All 6 false alarms occurred for the Natural condition. Further statistical analysis was not conducted because too little misses and false alarms occurred.

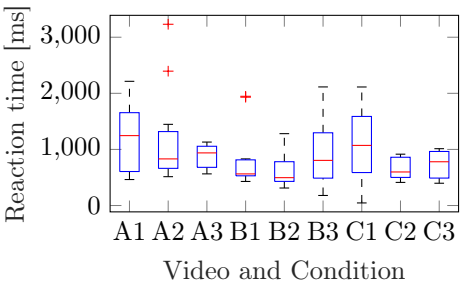


Figure 12: Reaction time for all possible combinations of video [1-3] and condition (A being Arbitrary, B being Semi-natural and C being Natural), for Exp 1.

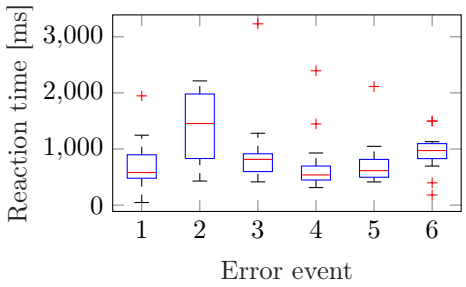


Figure 13: Reaction time for error event 1-6 in Exp 2. Two error events occurred per video: error event 1 and 2 occurred during Video 1, error event 3 and 4 occurred during Video 2 and error event 5 and 6 occurred during Video 3.

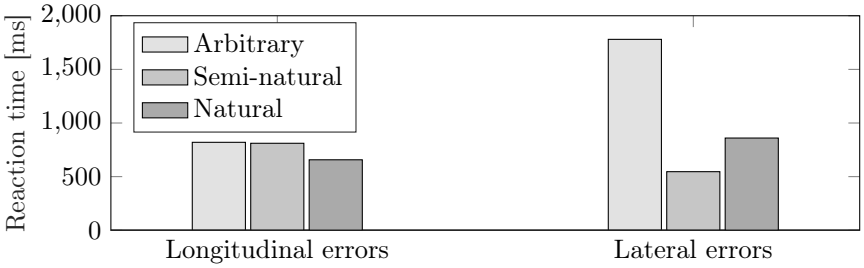


Figure 10: Average reaction time for three conditions, split by Longitudinal error events and Lateral error events, for Exp 2.

3.2 H2 Vigilance decrement

H2-2-a-i/ii Reaction time

Shown in Figure 14, the RT results from Exp. 1 are taken as the baseline. The dots depict the average RT of participants, per condition, for their first, second or third trial in Exp. 2. Each consist of 11 or 12 measurements, dependent on amount of misses, as two error events occur per video watched by 6 participants. RTs of responses ranged between 0 and 3480 ms. The amount of reactions with a RT over 2000 ms increased over time, as shown in Table 7. Because Exp. 1 and Exp. 2 include different people, different stimuli and different time intervals no further statistical test was conducted.

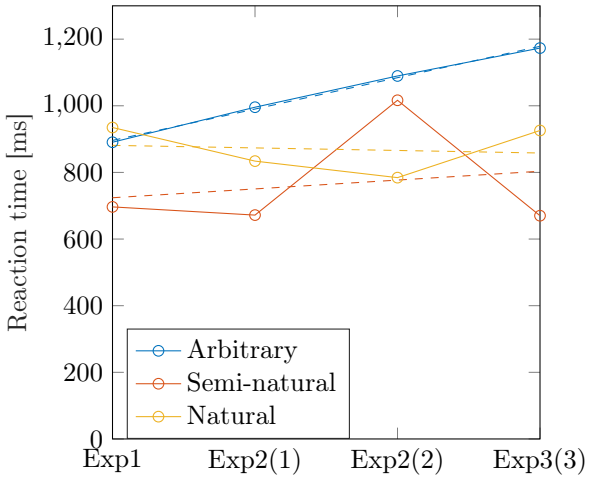


Figure 14: Reaction time at four consecutive moments, from baseline (Exp 1) to three trials in Exp 2. Dashed line depicts linear trend.

H2-2-a-i

We questioned if RT decrement over time is postponed dependent of the naturalism property of the interface. The trend lines in Figure 14 seem to suggest a correlation between steepness of the RT over time curve (which determines if, and how rapidly vigilance decrement occurs) and naturalism property, however these results should be interpret with caution, due to small sample sizes.

H2-2-a-ii

This hypothesis states that a more natural interface reduces RT decrement. This can not be concluded from the data. The trend lines in Figure 14 seem to suggest that the Arbitrary interface has the largest RT discrepancy over time, followed by the Semi-natural interface. The Natural interface shows no RT gain. Note that while this could possibly suggest a reduction, no constant levels of performance are maintained, thus over a longer timespan they could converge to the same RT performance level.

- Exp 1: (0 to 5.23 total task minutes), avg. RT to 17 error events, 5.23 minutes (54 video clips)
- Exp2(1): (5.23 to 10.23 total task minutes), avg. RT to 2 error events, 5 minutes (1 video clip)
- Exp2(2): (10.23 to 15.23 total task minutes), avg. RT to 2 error events, 5 minutes (1 video clip)
- Exp2(3): (15.23 to 20.23 total task minutes), avg. RT to 2 error events, 5 minutes (1 video clip)

H2-2-b-i/ii Accuracy

Only 2 misses occurred during Exp. 2: once during the first trial with the Arbitrary condition, and once during the third trial with the Natural condition. 6 False Alarms occurred, 3 of which were for one participant using the Natural interface condition, in Exp.

	Exp. 1	Exp. 2(1)	Exp. 2(2)	Exp. 2(3)
Amount of responses with RT >2000 ms	6	0	3	4
Total amount of responses	416	35	36	35
Percentage with RT >2000 ms	1.4 %	0.0 %	8.3 %	11.4 %

Table 7: Overview of response reaction time over 2000 ms

2(3). Because of the little amount of misses and false alarms no further statistical analysis was conducted.

H2-2-b-i

We hypothesised a delayed onset of decrement over time of accuracy. Too little data points were collected to interpret.

H2-2-b-ii

We hypothesised a reduced decrement over time of accuracy. Too little data points were collected to interpret.

3.3 H3 Self reported

H3-2-i Workload

The NASA TLX results are shown in Figure 15. The scores range from Very low (0) to Very high (100) for the mental demand, physical demand, temporal demand, effort, and frustration items, and from perfect (0) to failure (100) for the performance item. Error bars depict +1 and -1 SD. The Natural interface produced the lowest workload score in 6 out of 7 subdimensions (for the temporal demand it approximately tied with the other two conditions). A statistical significant effect was observed for the Mental Demand subdimension (printed in bold in Figure 15). No significant unequality was found for other subdimensions. The means of the three conditions for Mental Demand were unequal according to a one-way repeated measures ANOVA, $F(2,34) = 5.66$, $p = 0.008$. The ANOVA was followed by paired comparisons between the three conditions using a Bonferroni correction. Hence, in the pairwise comparisons, a result was declared statistically significant if the p value was smaller than 0.05/3. The pairwise comparisons indicated only one significant comparison: Subjects reported the task to be significantly ($p = 0.02$) more mentally demanding for the Semi-natural (M = 49.17) than for the Natural (M = 39.44) interface condition, with a 95% confidence interval of the difference between means from 1.31 to 19.62. The other comparisons were not significant, Arbitrary vs. Semi-natural: $p = 1.0$, Arbitrary vs. Natural: $p = 0.07$.

H3-2-ii Acceptance

Self reported results on Acceptance are shown on two axis, on Usefuless and Satisfying scale, in Figure 16. The scores range from very low (-2) to very high (2). Error bars depict +1 and -1 SD. The Natural condition is scored highest on both dimensions. No statistical significant effect is found, after a repeated measure one-way ANOVA was conducted for both scales. Usefulness: $F(2,34) = 2.38$, $p = .108$. Satisfying: $F(2,34) = 1.14$, $p = .331$. Full questionnaire results can be found in Appendix E.

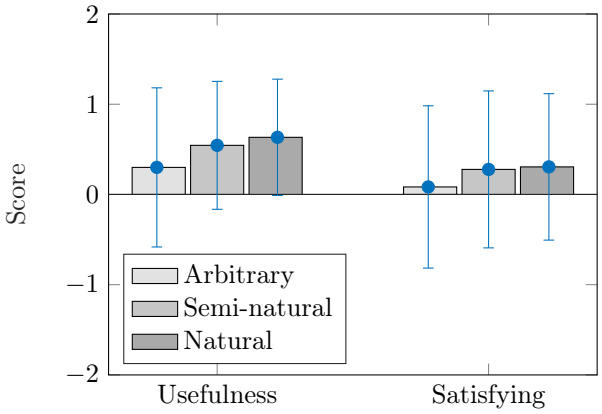


Figure 16: Self reported acceptance using the Van Der Laan acceptance questionnaire.

H3-2-iii Engagement

The engagement part of the DSSQ Short questionnaire was conducted, as shown in Figure 17. The scores range from 0 to 32. A repeated measure one-way ANOVA was conducted, which did prove the four conditions to be significantly unequal: $F(3,51) = 5.60$, $p = .002$. The ANOVA was followed by paired comparisons between the three conditions using a Bonferroni correction. Hence, in the pairwise comparisons, a result was declared statistically significant if the p value was smaller than 0.05/4. The pairwise comparisons indicated one significant comparison: Subjects reported that Pre-test (M = 22.2) was significantly ($p = .009$) more engaging than the Natural condition (M = 18.8), with a 95% confidence interval between means from .723 and 6.166. The other pairwise comparisons were not proven to be significant, Pre-test vs. Arbitrary: $p = 0.036$, Pre-test vs. Semi-natural:

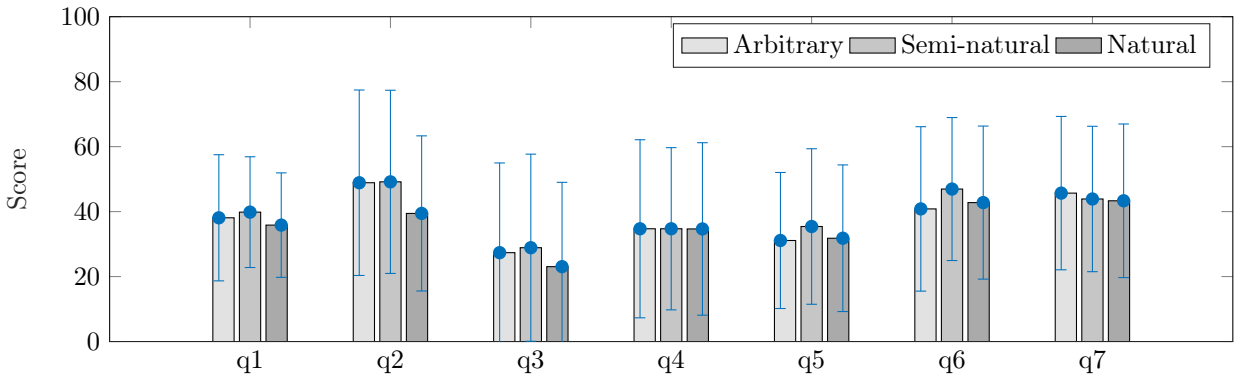


Figure 15: Self reported workload using Nasa TLX. Results were analysed using a repeated measure one-way ANOVA. Significant results are printed in bold.

	q1 Overall	q2 Mental	q3 Physical	q4 Temporal	q5 Performance	q6 Effort	q7 Frustration
$F(2,34)$	2.18	**5.66**	2.75	<0.01	0.88	1.09	0.19
p	0.129	**0.008**	0.078	0.999	0.426	0.346	0.828

$p = 0.022$, Pre-test vs. Natural: $p = 0.009$, Arbitrary vs. Semi-natural: $p = 1.0000$, Arbitrary vs. Natural: $p = 1.0000$, Semi-natural vs. Natural: p 1.000. Note that all pairwise comparisons between interface conditions result in ps of 1.000, which gives the impression that difference in engagement seems unlikely. Full questionnaire results can be found in Appendix F.

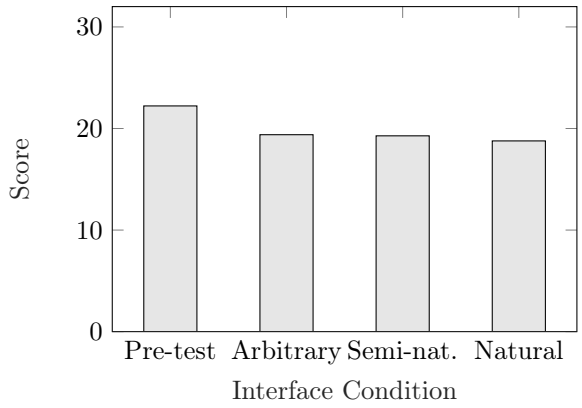


Figure 17: Self reported Engagement using the engagement part of the DSSQ Short questionnaire.

4 Discussion

The aim of this study was to examine the effects of naturalism on automation status monitoring of an automated vehicle. We analysed participant reactions to error events in various traffic situations with three different HMI's, in two exposure experiments.

4.1 Effects on reaction time and accuracy

Both Exp. 1 and 2 seem to suggest that the Semi-Natural condition results in the quickest RT and highest Hit rate. Exp. 1 shows the Arbitrary condition to be the next-best performing, while Exp. 2 shows it is the Natural condition. In this case the results from Exp. 2 are more trustworthy than Exp. 1. This is because in Exp. 2 conditions and error events are balanced, and in Exp. 1 they are not (Table 4 and 5). The difference in RT between error events can be as much as a factor 3 (Figure 9), thus balancing is critical.

When analysing the results from Exp. 2 further, it can be seen that there is a large difference per video 1 through 3, in Figure 10. While the Arbitrary condition always has the worst RT, the Natural condition does outperform the Semi-Natural condition for video 3.

Overall, the results seem to suggest that the Arbitrary interface results in the slowest RT, while both the Semi-natural and Natural conditions improve RT dependent of the error event in our experiment. Specifically for lateral automation status errors, the Arbitrary interface performs approximately twice as slow as the other conditions, as shown in Figure 10.

We believe this large difference between the Arbitrary interface and the Semi-Natural interface, and to some extent the Natural interface, is due to the fact that latter two interfaces did respond dynamically to precipitating cues (e.g. dynamic curvature of the stripes due to a curved road, or movement of the dot or ellipse due to an object being detected, examples in Figure 5). This indicated to the participant when to be alert, resulting in faster RTs. The Arbitrary interface was completely static. These dynamic movements indicating a higher error event probability were mentioned multiple times by participants during debrief.

We are inclined to believe that the Semi-Natural interface outperformed the Natural interface is because the Natural interface was more difficult to interpret at first. This was also mentioned multiple times during debrief. This seems plausible because the RT Natural interface decreased from an average of 935 ms to 759 ms between Exp. 1 to Exp. 2 (Figure 6 and 7). For our experiment, no training was done and explanation of interfaces were still images. In further research, we recommend some form of training or more in-depth explanation before the Exposure Experiments.

Data measurement regarding accuracy performance might not have been sensitive enough. Based on the data, no conclusions are drawn. For future studies we recommend to improve data measurement sensitivity, perhaps through a different measure of performance error, or with added distractors or tasks.

4.2 Effects on vigilance decrement

The amount of responses with a high RT (>2.0s) increases over time, as can be seen in Table 7, which gives the impression some form of vigilance decrement occurred. Also, the eventrate decreased between Exp. 1 and 2, and the percentage of responses a high RT increases from 1.7 % to 6.6 %. This is in line with Hancock, 2015: "If you build vehicles where drivers are rarely required to respond, then they will rarely respond when required." [6].

The trendlines in Figure 12 seem to support H2-2-a-i and H-2-2-a-ii. Both the onset and decrement of RT seem influenced by naturalism. The trendlines show the largest RT increase over time for the Arbitrary condition, less RT increase over time for the Semi-Natural condition and even an decrease of RT over time for the Natural condition. While these results give the impression that the Vigilance decrement is improved when using a Natural interface, these results should be interpreted with caution due to small sample sizes.

4.3 Effects on self-reported workload

The data suggests that mental demand of the participants was influenced by which interface condition was used. The Natural condition was scored to result in

approximately 10 percentage points less mental demand than the Arbitrary and Semi-Natural conditions, and the three interface conditions proved significantly unequal. As mental demand is the subitem most tied to the presumed cognitive mechanism, these results hint that natural information, precipitating cues or decreased cognitive switching could have reduced cognitive strain due to some form of externally stored situation awareness or active externalism.

4.4 Effects on self-reported acceptance

Descriptive results seem to suggest that more natural HMIs are perceived more useful and satisfying, however, these results did not prove to be significant due to large variance. For future research we do believe this measurement to be of relevance, and additional questions could further clarify the underlying motivation (e.g. aesthetics or automation trust).

4.5 Limitations

Limitations by this virtual simulation experiment are acknowledged. First of all, a part of the presumed cognitive mechanism was that the Natural interface decreases mental effort of cognitive switching. While the results show a decreased Mental Demand for the Natural condition, it is unclear whether participants actually were switching between driving environment and interface. An effort was made to decrease likeliness of "Pressing the button when the blue interface disapears" by adding a cognitive step in choosing whether the error is Longitudinal or Lateral, and by asking the participants to also respond by steering and possibly braking. It is unknown whether these efforts did affect in cognitive switching. We recommend further research with eye-tracking.
Second, the studies set out to compare Arbitrary and Natural interfaces. While the interface designs are made to include Arbitrary and Natural interface properties, these may not have been flawlessly designed. It is hard to distinguish the effects due to this specific interface, and due to naturalism. Future research with eye-tracking could expose these subtle effects due to interface design. Also, future research can include different interfaces and type of information.
Third, while the use of real-life driving videos implies realistic visual image and vehicle dynamics, the simulator set-up is of low physical fidelity, as it lacks real cockpit geometry, motion cues, and both haptic and auditory feedback. We did not ask participants about perceptual fidelity. Also, participants knew they could not be physically harmed by a potential crash and were probably less stressed. Lack of evidence of behavioural fidelity means these results should be interpreted with caution, and a real world replication of this experiment is advised.
Fourth, the results show little usable results in accuracy as little misses and false alarms occurred. A more challenging task design or a secondary task could improve this. Fifth, some parts of the experiment set-up could have been more specific. Participants were free to keep their hands where they wanted, mostly on the steering wheel or on their lap. This has influence on RT. Sixth, Participants were students between 20 and 28 years old, while a wider demographic drives. The student participants did indicate that they were motivated in the pre-test questionnaire. Lastly, using more participants or designing the experiment to include more responses is advised to increase statistical evidence.

5 Conclusion

Due to rapid development of automated vehicles, humans will increasingly be in a supervisory role, posing human factors challenges to maintain safety and comfort. Visual augmented reality is a technology that could help overcome these challenges. We introduced a scale to classify types of AR from Arbitrary to Natural, and we designed and tested three HMI interfaces using real life driving videos in a simulation experiment. Results seem to suggest that, in comparison to an Arbitrary interface, the Semi-Natural interface could decrease reaction times to automation status Error Events. The Natural interface could also decrease reaction times, but to a lesser extent, and could also decrease mental demand. From the results it seems that a more Natural interface could potentially decrease vigilance decrement compared to an Arbitrary interface, but these results are preliminary and require further research. To review, we believe more Natural interfaces show potential to improve safety and comfort when becoming technologically feasible if further human factors research is conducted.

References

[1] Gold C, Damböck D, Lorenz L, Bengler K. "Take over!" How long does it take to get the driver back into the loop? Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2013;57(1):1938–1942.

[2] Kyriakidis M, Happee R, De Winter JCF. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. Transportation Research Part F: Traffic Psychology and Behaviour. 2015;32:127–140.

[3] Parasuraman R, Riley V. Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors. 1997;39(2):230 – 253.

[4] SAE. Automated Driving: Levels Of Driving Automation Are Defined In New SAE International Standard J3016. Global Ground Vehicle Standards. 2016;.

[5] Bainbridge L. Ironies of Automation *. 1983;19(6).

[6] Hancock PA. Automobility: The coming use of fully-automated on-road vehicles. 2015 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision, CogSIMA 2015. 2015;p. 137–139.

[7] U S Department of Transportation. Automated driving systems 2.0: A vision for safety;.

[8] Kyriakidis M, Winter JCFD, Stanton N, Bellet T, Arem BV, Martens MH, et al. Theoretical Issues in Ergonomics Science A human factors perspective on automated driving. Theoretical Issues in Ergonomics Science. 2017;0(0):1–27.

[9] Mackworth NH. The breakdown of vigilance durning prolonged visual search. Quarterly Journal of Experimental Psychology. 1948;1(1):6–21.

[10] Warm JS, Parasuraman R, Matthews G. Vigilance Requires Hard Mental Work and Is Stressful. Human Factors. 2008;50(3):433–441.

[11] Davies DR, Parasuraman R. The psychology of vigilance. Academic Pr; 1982.

[12] Cabrall CDD, Happee R, De Winter JCF. From Mackworth's clock to the open road: A literature review on driver vigilance task operationalization. Transportation Research Part F: Traffic Psychology and Behaviour. 2016;40:169–189.

[13] Patricia L, Susan L. A Novel Active Heads-Up Display for Driver Assistance. IEEE transactions on systems, man, and cybernetics Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society. 2009;39(1):85 – 93.

[14] Dokic, j; Muller, B; Meyer G. European Roadmap Smart Systems for Automated Driving. European Technology Platform on Smart Systems Integration. 2015;p. 1–39.

[15] Azuma R. A survey of augmented reality. Presence: Teleoperators and Virtual Environments. 1997;6(4):355–385.

[16] Azuma R, Baillot Y, Feiner S, Julier S, Behringer R, Macintyre B. Recent Advances in. Ieee Computer Graphics And Applications. 2001;21.6(December):34–47.

[17] Gaver W. Technology affordances. Acm. 1991;p. 79–84.

[18] Milgram P, Kishino F. Taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems. 1994;E77-D(12):1321–1329.

[19] Chiappe D, Strybel TZ, Vu KPL. A Situated Approach to the Understanding of Dynamic Situations. Journal of Cognitive Engineering and Decision Making. 2015;9(1):33–43.

[20] Clark A, Chalmers D. The Extended Mind. Analysis. 1998;58(1):7–19.

[21] Van Der Laan JD, Heino A, De Waard D. A simple procedure for the assessment of acceptance of advanced transport telematics; 1996.

[22] Helton WS. Validation of a Short Stress State Questionnaire. In: Proceedings of the human factors and ergonomics society 48th annual meeting; 2004. p. 1238–1242.

6 Appendices

6.1 Appendix A: Literature review

A literature review on automotive augmented reality

By

J. T. Pijnenburg

4103157

Abstract

This literature study surveys the field of automotive augmented reality. Augmented reality, in which virtual objects are superimposed or composited with the real world, could have the potential to improve visual automotive human-machine interfaces. To discuss interface properties that enhance reality in a lively, intuitive way, in contrast to arbitrary icons, we introduce the term naturalism. This scale ranges from arbitrary to natural. This study describes the current state of automotive augmented reality on market and in scholarly publications, and how presented interfaces relate to naturalism, by consulting approximately 50 publications. The summarised body of knowledge is described below.

Many automobile producers offer vehicles with a head-up display. As this enhances reality using symbolic information, we argue that this is can be classified as a type of arbitrary AR. Automobile producers show intention of producing more natural interfaces as they showcase their newest concept cars. In scholarly publications, a wide array of use cases for automotive augmented reality have been researched, including eyes-on-road information, hazard and pedestrian warning, take over request enhancement, and more. In three papers, test warning systems in which conditions relate closely to naturalism. While it is early to draw conclusions, it seems that arbitrary interfaces are fit to convey simple, clear messages, and natural interfaces are fit for offering more rich information such as positional information of a pedestrian. We believe different use cases for automotive augmented reality could benefit from research regarding naturalism, to improve human-machine interaction.

TU Delft

Contents

1 Introduction

Augmented reality (AR) is defined as a variation of virtual reality, in which the user can see the real world with virtual objects superimposed or composited with the real world[1 2]. AR could have the potential to improve a visual human-machine interface by exploiting natural cues, as it can be used to highlight both hidden affordances and perceptible affordances[3]. The user receives focused information enabling him to act upon this affordance. Typically, adding this information involves using computer-generated graphics so they appear in the real world[4]. The use of AR in human factors has been a subject of research for about four decades, leveraging human-machine interaction to improve safety in aviation[5 6 7 8].

1.1 AR in Automotive

Use of AR also shows potential in automotive applications, as AR devices have potential to improve driving performance[9]. The intention is to use this technology to positively affect the human-machine interaction, however one could argue that AR could negatively affect this, e.g. leading to information clutter or information overload. In their European roadmap Dokic et. al. predict automotive AR to be in the research and development phase untill 2020, then transitioning to a demo phase and hitting product and market from 2022[10].

1.2 Naturalism of an AR interface

In contrast to the dashboard, which informs the user by showing arbitrary icons only in a virtual environment, an AR interface can also emphasize reality. Is the opportunity of informing the user in a vivid way worthwhile? Does highlighting actual real life cues propose any advantages? To be able to discuss these questions we introduce a new property to describe this quality of an AR interface. We name this interface property 'Naturalism', as a natural interface reveals real-life clues, using realistic motion, real depth-perception and intuitive stimuli. In contrast, an arbitrary AR interface relies on knowledge-based icons, numbers and gauges (Figure 1).
The term Naturalism was chosen as the preferred terminology. While not perfect it seems sufficiently fitting for now as it is somewhat self-explanatory and emphasizes the contrast between arbitrary and natural properties. The main criteria were self-explanatory value and potential for confusion with existing concepts. The term 'level of augmentation', also has self-explanatory merits but is too easily misunderstood as this term could also relate to Milgram's Virtuality Continuum, which runs from real environment to virtual environment, with AR inbetween[11]. The term 'skeuomorphism' was also considered, but the definition related too much to using old properties of previous design solutions. Old design properties help intuitive use in a cultural, knowledge-based way, which doesn't contrast with static arbitrary icons the way a natural intuitive design does, thus this term was found insufficiently fitting.

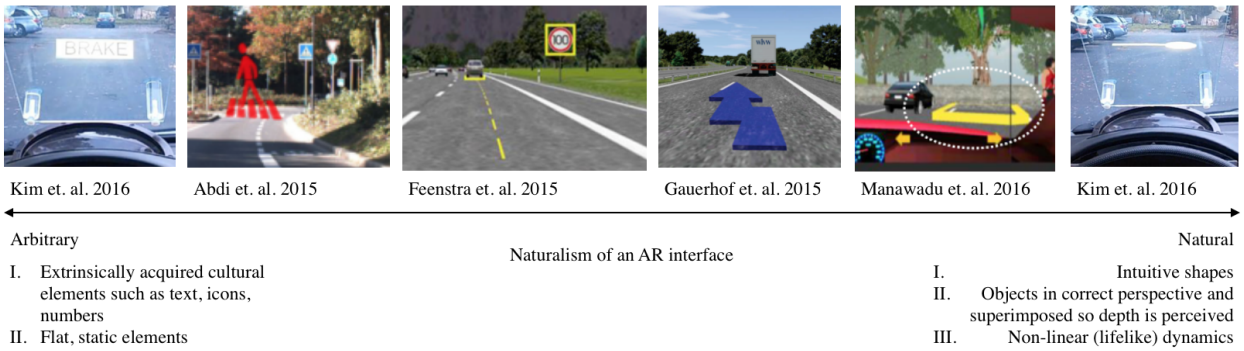


Figure 1: Naturalism scale

2 Materials and methods

The goal of this literature review is to form an overview of the current state of research on the use of AR in automotive applications, including current issues and design. My research included the following questions.

1. What is the current state of automotive AR on market and in development, and how natural are these interfaces?

2. What is the current state of automotive AR in scholarly publications, and how natural are these interfaces?

For question 1, automobile producer websites and car manuals were conducted, as well as automotive and design news sources (e.g. designboom and carbodydesign).

Next a systematic search was conducted on scholarly publications in Google Scholar. The following keywords were used: "Augmented Reality" AND "Human Factors" AND "Driver" AND "automated" OR "automation" AND "automotive" OR "car" OR "Vehicle". Only English papers were included. From 629 results, 63 were selected based on title. Next, we went through cited references, ultimately selecting relevant papers from a pool of 135 documents.

3 Currently on market and in development

Various types of automotive AR interfaces are available on market currently, which will be discussed in this chapter. By far the most common application is the head-up display (HUD), which shows different types of information on the bottom side of the windscreen, in vision of the driver. We also discuss AR for rear-view camera applications. Information on acoustic AR was also found, but it was decided to focus solely on visual AR for this review.

3.1 Head-up Display

In 2016 releases, many on-market automobiles include HUDs. Most are projected on the windscreen, showing monochrome or bicolor projections of information on speed and navigation. In others, information is projected on a small transparent shape in front of the windscreen, that can fold down when not in use. We argue that these HUDs are a type of arbitrary AR. Virtual objects are composited with the real world, showing properties of reality (e.g. your speed compared to the environment) and otherwise hidden information (e.g. navigational clues). This information is being shown using text, icons and numbers, which are static and flat, thus we categorise it on the most arbitrary side of the naturalism spectrum. Table 1 shows a non-exhaustive overview of AR-HUD interfaces for 2016 automobile releases.

	Speed*	Navigation	Gear	Adaptive cruise control	Phone calls	Music	Lane keeping assist	Hybrid drivetrain status
2016 Lexus RX	x	x	x	x			x	
2016 Toyota Prius	x	x		x			x	x
2016 Volvo XC90	x	x		x	x			
2016 Mini Cooper	x	x			x	x		
2016 Corvette Stingray	x		x					
2016 Jaguar XF	x	x	x					
2016 Mercedes-Benz C-class	x	x						
2016 BMW 7 series	x	x						
2016 Mazda Mazda3	x	x						
2016 Ford Shelby GT350			x					

Table 1: Information shown in AR-HUD for select 2016 cars, non-exhaustive. These interfaces show information in a arbitrary manner. *Can also include maximum speed and RPM

3.2 AR for rear-camera driving

AR is also being used as a driving aid when driving backwards[12]. An example of this is the Delphi Parking Guidance System (Figure 2), produced by automotive parts supplier Delphi, which features both static and dynamic on-screen overlays to support the driver in parking.

Figure 2: The Delphi Parking Guidance System superimposes guidelines on the vehicle display, courtesy of Delphi Automotive LLP

3.3 In development

Optical see-through AR on the whole front windscreen has currently not been fully developed for commercial consumption[12]. Most Automotive companies do show intentions to develop AR as they showcase in their newest concept cars. A recent example is the BMW Vision Next 100, in which BMW tries to improve information availability, performance and safety trough a head-up display on the front windscreen (Figure 3). In 2015, Mini showed the Augmented Vision Glasses[13] concept which uses AR to look at a dog in the blind spot, through the car (Figure 4). Both concepts show naturalistic interface properties. It can be concluded that all applications currently on market show arbitrary information, however, the car makers' design studios release future concepts showcasing more naturalistic interfaces.

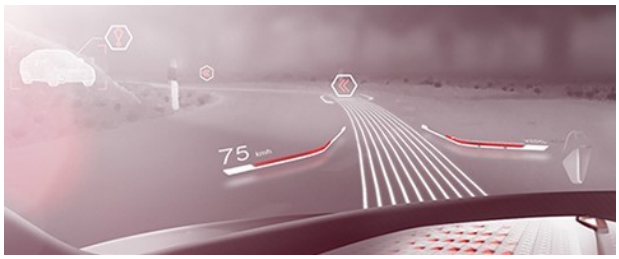


Figure 3: BMW Vision Next 100 interior sketch by H.v. Freyberg from BMW Design

Figure 4: MINI Augmented Vision: A revolutionary display concept offering enhanced comfort and safety.

4 In Scholarly Publications

Already in 1995, automotive usecases for AR HUDs were a topic of research, including conveying information about cellular telephones, navigation systems, vehicle-to-roadside communcation systems and others.[14] In this chapter an overview of current use-cases discussed in scholarly publications is given, and their relation to human factors. Next, potential drawbacks of using AR are discussed. Lastly, we discuss AR interfaces in literature with naturalistic properties.

4.1 AR use-cases in publications

4.1.1 Eyes-on-road information

A much proposed hypothesis regarding the advantage of an AR-HUD compared to a central information display is the spatial location of an AR-HUD being closer to the focus point of the driver, as they observe the road, which doesn't require them to gaze down, avoiding attention gaps[15 16 17 18 19 9 20]. However, it is also argued that AR and HUD could supplement dashboard screens, but will not replace them since different types of information is displayed[21]. As in-vehicle information systems distract[22 23], if the right information display split is made[14], this can decrease the distraction the head unit interface poses.

4.1.2 Hazard and pedestrian warning

Multiple studies found positive effects when using an AR system to warn the driver about hazards and pedestrians, including increased response rate[24], decreased collisions[25], smoother braking behaviour[26], improved driver cognition[27] and increased hazard detection likelihood[28]. Two studies specifically noted

there were no significant negative costs related to the AR presence (e.g. interference)[24][28].

Another interesting consideration was found using a high fidelity driving simulator, in which both young and old drivers tested an AR system for hazard avoidance. For both age groups, results showed that it can decrease collisions, especially to help older drivers to avoid sudden hazards. An interesting effect occured: while older drivers became more careful, younger drivers seemed to rely on the AR without adopting safer driving behaviour[25]. It seems that young drivers develop trust more quickly, implying the need for highly reliable systems. An alternative approach is to design a system that is transparent about its reliability, so humans can match their trust thus decreasing the likeliness of over-reliance.

4.1.3 Take over request

Researchers at the BMW Group Research and Technology[29], show that AR support during take-over scenarios does not influence take over times, but does affect quality of reaction. They do so by testing two AR concepts against a control condition. "AR red" provides information on where the hazard is that caused the take over request (TOR), e.g. the accident location, while "AR green" provides information on where it is safe to steer to. During the 7-second TOR window, AR green shows most improvements in steering actions. The writers will conduct further research regarding combinations of both concepts.
AR may also improve situational awareness during a TOR, as demonstrated by Laglois et. al.[30]. They hypothesise that superimposition of the information onto the real spatiotemporal context, which AR allows, helps the driver detect, analyse and react appropriately. Thus improving situational awareness, and adoption of safer driving behaviour. In their study, the participant has to execute a lane change manoeuvre with one of three conditions. The baseline is manual driving. Next is a transition from automated driving to manual driving, using a HUD. The third condition is the same, but using an AR-HUD. With AR, the drivers' manoeuvre and brake pedal handling improve, on which the writers base their assumption that situational awareness improved, proving their hypothesis to be true.

4.1.4 Cooperative awareness

Introducing cooperative awareness, Gomes et. al. show an use-case where vehicle-to-vehicle communication of camera data and AR are leveraged to build an interface that allows users to see through the vehicle in front[31][32]. After testing a prototype they conclude that this system does improve the safety of overtaking large and vision-obstructing vehicles.

4.1.5 Communication

Gauerhof et. al. regard an AR system not based on the windscreen, but in wearable glasses, which is portable and can be retrofitted. Concluded is that an AR-application is not a sufficient condition when improving communication to automated cars. Participants do score the system high on trust scale[33].

4.1.6 Virtual traffic lights

An interesting use of AR is to replace traffic lights, by adding them to a virtual interface. This could improve traffic flow and reduce collisions. Olaverri-Monreal et. al. found that using virtual traffic lights had no significant influence on driver performance, which could be promising if virtual traffic lights were to replace physical infrastructure[34]. If further research proves that no safety issues occur, potential upsides for this technology can be investigated. They did not however research other forms of interfaces, so it could very well be that this technology works just as well when the information is not provided using AR.

4.1.7 Motion sickness

An unexpected possible upside of using AR can be reduction of motion sickness, as shown in an aircraft simulator experiment[35]. While this has potential to avoid the visual-vestibular conflict, it could introduce sensory conflicts when the visuals are dynamic and the velocity constant or when static visuals are shown during acceleration and braking (e.g. rush hour traffic)[36].

4.1.8 Mindfulness

Paredes et. al. explore the opportunities for mindfulness, using AR/VR in an automated vehicle to im-

merse users in soothing scenarios[37].

4.1.9 Effect on human factors

In this section a wide range of AR use-cases in current publications was shown. We conclude that, while various human factors effects are tested in these studies, there seems to be a similarity. Many of the AR Interfaces did not help significantly improving performance (e.g. quick braking reaction time), but did help to improve the quality of the interaction, e.g. a smoother braking behaviour. This could potentially suggest that AR is best fitted not to improve human perception threshold, but rather to improve the cognitive processing, not leading to quicker reactions but more appropriate interaction.

4.2 Pitfalls of AR

HUDs in windshields also pose a risk, as they coud block or distort information[12]. Using in-vehicle VR during a pilot, three out of eight partifipants experienced motion sickness[37].

AR cues can also distract and overload the perceptual abilities of a driver, which can lead to trust decrement (e.g. ignoring the cues)[38][39]. Also, high false alarm rates caused by poor system reliability may also affect user trust[40][41]. When users can't trust AR warnings this leads to a decreased performance[42][43], thus designers need to prioritise system reliability.

4.3 Naturalism of AR in literature

In scholarly publications, a wide array of AR interfaces are used. Interfaces found in publications previously discussed, are shown in figure 5, and ranked along the naturalism continuum. There is a large variation in naturalism of these interfaces.
In three papers specific differences between natural and arbitrary interfaces were discussed. All three studies were on types of warning systems, two on pedestrian detection and one on speed limit. In both of the pedestrian detection cases, the more natural interface performed better, due to superimpositioning properties. In the speed limit case the natural interface does not perform better, but note that a speed warning doesn't require any positional information. While it is early to draw conclusions, from these three studies, it seems that when positional information helps the driver, the superimpositioning property of a natural interface is of great influence. When less rich information is required to interpret the warning, a naturalistic interface is of less help. These three studies are discussed in this chapter.

4.3.1 Pedestrian detection

After an experiment improving pedestrian detection using AR, Phan et. al. asked participants which cue they preferred. 17 of 25 participants chose the bounding box cue (a cue sticking to the pedestrian, more natural) over the pedestrian panel (a static warning sign, more arbitrary), argueing that the bounding box directly showed where the pedestrian was while the pedestrian panel only forced them to brake and make them distracted[51].
Kim. et. al. also tested a pedestrian warning system, which participants tested for two conditions[26]. Either a warning with the text 'Brake' (arbitrary) was given, or a 'Virtual Shadow' was applied to the pedestrian (natural). Besides also giving positional information, they argued that this virtual shadow was at the same focus distance as the pedestrian, not causing the driver to shift focus as was needed for the arbitrary depthless sign. Results shows that both interfaces are an improvement compared to the baseline in reaction time, but the natural interface shows much smoother braking behaviour.

One aspect of a more naturalistic interface is real-world fixated information, in constrast to static information. These allow the user to simultaneously process augmented and real information[53].

Figure 5: Overview of automotive AR interfaces shown in scholarly publications

Table 2: Authors of Figure 5

26	44	9	45	46
27		47	33	31
48	29	33	49	
25	17	50	27	
17	51		49	30
17	44		34	52

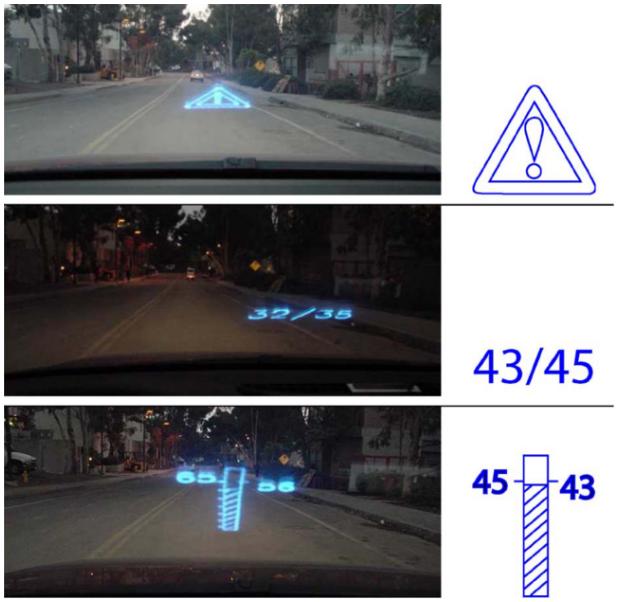


Figure 6: Speed warning interfaces by Patricia et. al.[20], ranging from most arbitrary (top) to a bit more natural (bottom)

4.3.2 Speed warning

Another example of a study in which naturalism plays a role is an impressive AR study conducted on real roads[20] (Figure 6). The researchers use three different interfaces to inform the driver of speed limits and compare these to a conventional speedometer. A laser is used to project these interfaces on a windscreen. One interface is a triangular exclamation point warning (arbitrary). The second interface shows numbers of the current speed and the road speed limit (e.g. 37/50, arbitrary). The third interface shows a bar of which percentage is filled corresponding to the current speed divided by the maximum speed (more natural). The third interface also includes some smaller numbers, and the graphic shows a bouncing motion when the speed limit is reached. These interfaces were judged on how quickly the driver would slow down when warned. All three interfaces outperformed the base condition, but the most arbitrary warning- the exclamation mark in a triangle- worked best. This can be attributed to the fact that this was only displayed when going over the speed limit, thus catching a lot of the driver's attention. The arbitrary warning sign was also easy to comprehend; it is a much used icon. The most natural interface was a bit more complex to comprehend, but did include more information. It seems to be a trade-off between wanting to catch attention when warning, but keeping the driver's eyes on the road as much as possible, and at the same time a trade-off between interfaces that are easily understood but also rich in information.

5 Conclusions and recommendations

We introduce the naturalism dimension to discuss arbitrary and more natural properties of an AR interface for automotive applications. While this field is still in its infancy, we are able to draw some preliminary conclusions about the current state of the market, in scholarly publications and on design methodologies. Automobile producers currently offer AR systems, which are all very arbitrary. Their design concepts show plans for more naturalistic interfaces when technology allows fully augmented systems. Scholarly publications seem to suggest that AR is best fitted not to improve human perception threshold, but rather to improve the cognitive processing, not leading to quicker reactions but more appropriate interaction. Arbitrary interfaces seem to excel at showing clear warning signals, while natural interfaces seem less suitable for simple warnings, but seem to provide rich information regarding a more complex warning, e.g. positional information of a pedestrian.
We conclude that from both on market and in publications, a variety of interfaces along the naturalism dimension can be found. It seems, while preliminary, that there are potential advantages and disadvantages to using a more natural interface. More human factors research regarding naturalism seems meaningful and could improve automotive human-machine interaction more safe and comfortable.

References

[1] Azuma R. A survey of augmented reality. Presence: Teleoperators and Virtual Environments. 1997;6(4):355–385.

[2] Azuma R, Baillot Y, Feiner S, Julier S, Behringer R, Macintyre B. Recent Advances in. Ieee Computer Graphics And Applications. 2001;21.6(December):34–47.

[3] Gaver W. Technology affordances. Acm. 1991;p. 79–84.

[4] Kearney JK, Rizzo M, Severson J. Virtual reality and neuroergonomics. Neuroergonomics: The brain at work. 2007;p. 253–274.

[5] Fischer E, Haines RF, Price TA. Cognitive Issues in Head-Up Displays. NASA CR-3137. 1980;.

[6] Fischer. The Role of Cognitive Switching in Head-Up Displays; 1979.

[7] Naish JM. Combination of Information in Superimposed Visual Fields. Nature. 1964;202(No. 4933):641–646.

[8] Naish JM, Miller DL. An Experimental Evaluation of Head-Up Display Formats; 1980. July.

[9] Medenica Z, Kun AL, Paek T, Palinko O. Augmented reality vs. street views: a driving simulator study comparing two emerging navigation aids. Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '11). 2011;p. 265–274.

[10] Dokic, j; Muller, B; Meyer G. European Roadmap Smart Systems for Automated Driving. European Technology Platform on Smart Systems Integration. 2015;p. 1–39.

[11] Milgram P, Kishino F. Taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems. 1994;E77-D(12):1321–1329.

[12] Gabbard JL, Fitch GM, Kim H. Behind the glass: Driver challenges and opportunities for AR automotive applications. Proceedings of the IEEE. 2014;102(2):124–136.

[13] BMW Group Innovation and Design Communications. MINI Augmented Vision: A revolutionary display concept offering enhanced comfort and safety.; 2015.

[14] Beyerlein DG. Head-Up Displays of the Future. SAE Transactions. 1995;(Paper 95096):950960.

[15] Green P. The 15-second rule for driver information systems. In: Proc of the ITS America. Cd; 1999. p. 1–9.

[16] Dingus TA, Hulse MC, Antin JF, Wierwille WW. Attentional demand requirements of an automobile moving-map navigation system. Transportation Research Part A: General. 1989;23(4):301–315.

[17] Abdi L, Abdallah FB, Meddeb A. In-Vehicle Augmented Reality Traffic Information System: A New Type of Communication between Driver and Vehicle. Procedia Computer Science. 2015;73(Awict):242–249.

[18] Haeuslschmid R, Pfleging B, Alt F. A Design Space to Support the Development of Windshield Applications for the Car. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016;p. 5076–5091.

[19] Kerschbaum P, Lorenz L, Hergeth S, Bengler K. Designing the human-machine interface for highly automated cars - Challenges, exemplary concepts and studies. Proceedings of IEEE Workshop on Advanced Robotics and its Social Impacts, ARSO. 2016;2016-March.

[20] Patricia L, Susan L. A Novel Active Heads-Up Display for Driver Assistance. IEEE transactions on systems, man, and cybernetics Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society. 2009;39(1):85 – 93.

[21] Bennakhi A, Safar M. Ambient Technology in Vehicles: The Benefits and Risks. Procedia Computer Science. 2016;83(1):1056–1063.

[22] Rümelin S, Siegl P, Butz A. Could You Please . . . ? Investigating Cooperation In The Car. In: Adjunct Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 13),(October 28th-30th 2013), Eindhoven, The Netherlands;. p. 61–64.

[23] Harvey C, Stanton NA, Pickering CA, Mcdonald M, Harvey C, Stanton NA, et al. To twist or poke ? A method for identifying usability issues with the rotary controller and touch screen for control of in-vehicle information systems. Ergonomics. 2011;54(No. 7, July 2011):609 – 625.

[24] Rusch ML, Schall MC, Gavin P, Lee JD, Dawson JD, Vecera S, et al. Directing driver attention with augmented reality cues. Transportation Research Part F: Traffic Psychology and Behaviour. 2013;16(319):127–137.

[25] Fu WT, Gasper J, Kim SW. Effects of an in-car augmented reality system on improving safety of younger and older drivers. 2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013. 2013;p. 59–66.

[26] Kim H, Miranda Anon A, Misu T, Li N, Tawari A, Fujimura K. Look at Me. Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16. 2016;(March):294–298.

[27] Park Bj, Lee Jw, Yoon C, Kim Kh. A Study on Augmented Reality of a Vehicle Information using Head-Up Display Augmented Reality Information in a vehicle. 2016;.

[28] Schall MC, Rusch ML, Lee JD, Dawson JD, Thomas G, Aksan N, et al. Augmented Reality Cues and Elderly Driver Hazard Perception. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2012;55(3):643–658.

[29] Lorenz L, Kerschbaum P, Schumann J. Designing take over scenarios for automated driving : How does augmented reality support the driver to get back into the loop? Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting - 2014. 2014;1:1681–1685.

[30] Langlois S, Soualmi B. Augmented reality versus classical HUD to take over from automated driving: an aid to smooth reactions and to anticipate maneuvers*. 2016;33(0):1571–1578.

[31] Gomes P, Vieira F, Ferreira M. The See-Through System : From Implementation to;.

[32] Pedro Gomes, Cristina Olaverri-Monreal MF. Making Vehicles Transparent Through V2V Video Streaming. IEEE Transactions on Geoscience and Remote Sensing. 2004;p. 1–11.

[33] Gauerhof L, Alexander K, Lienkamp M. ADAS for the Communication between Automated and Manually Driven Cars. 7 Tagung Fahrerassistenz. 2015;p. 8.

[34] Olaverri-Monreal C, Gomes P, Silveria MK, Ferreira M. In-Vehicle Virtual Traffic Lights: A graphical user interface. Information Systems and Technologies (CISTI), 2012 7th Iberian Conference on. 2012;(December 2015):1–6.

[35] Feenstra PJ, Bos JE, Van Gent RNHW. A visual display enhancing comfort by counteracting airsickness. Displays. 2011;32(4):194–200.

[36] Diels C, Bos JE. Self-driving carsickness. Applied Ergonomics. 2016;53(October):374–382.

[37] Paredes P, Ju W, Landay J. The Mindful Commute – Cars as Intervention Machines;.

[38] Schall Jr M, Rusch M, Lee J, Vecera S, Rizzo M. Attraction without distraction: Effects of augmented reality cues on driver hazard perception. Journal of Vision. 2010;10(7):236–236.

[39] Rusch ML, Schall MC, Lee JD, Dawson JD, Rizzo M. Augmented reality cues to assist older drivers with gap estimation for left-turns. Accident Analysis and Prevention. 2014;71(319):210–221.

[40] Sorkin RD. Why are people turning off our alarms? Journal of acoustic society of america. 1988;84(3):1107–1108.

[41] Bliss JP, To. Alarm Reaction Patterns by Pilots as a Function of Reaction Modality. International Journal of Aviation Psychology. 1997;7(1):1–14.

[42] Bliss JP, Acton SA. Alarm mistrust in automobiles: How collision alarm reliability affects driving. Applied Ergonomics. 2003;34(6):499–509.

[43] Maltz M, Shinar D. Imperfect In-Vehicle Collision Avoidance Warning Systems Can Aid Drivers. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2004;46(2):357–366.

[44] Kim H, Isleib JD, Gabbard JL. Casting shadows: Ecological interface design for augmented reality pedestrian collision warning. Proceedings - IEEE Virtual Reality. 2016;2016-July:205–206.

[45] Lee JW, Yoon CR, Kang J, Park BJ, Kim KH. Development of lane-level guidance service in vehicle augmented reality system. International Conference on Advanced Communication Technology, ICACT. 2015;2015-Augus:263–266.

[46] Manawadu UE, Kamezaki M, Ishikawa M, Kawano T, Sugano S. A haptic feedback driver-vehicle interface for controlling lateral and longitudinal motions of autonomous vehicles. IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM. 2016;2016-Septe(July):119–124.

[47] Zimmermann M, Bauer S, Lutteken N, Rothkirch IM, Bengler KJ. Acting together by mutual control: Evaluation of a multimodal interaction concept for cooperative driving. 2014 International Conference on Collaboration Technologies and Systems, CTS 2014. 2014;(July):227–235.

[48] Zimmermann M, Bengler K. A multimodal interaction concept for cooperative driving. IEEE Intelligent Vehicles Symposium, Proceedings. 2013;(July 2015):1285–1290.

[49] Pfannmüller L, Kramer M, Senner B, Bengler K. A comparison of display concepts for a navigation system in an automotive contact analog head-up display. Procedia Manufacturing. 2015;00(Ahfe):0–0.

[50] Damböck D, Weißgerber T, Kienle M, Bengler K. Evaluation of a Contact Analog Head-Up Display for Highly Automated Driving. 4th International Conference on Applied Human Factors and Ergonomics. 2012;p. 6011–6020.

[51] Phan MT, Thouvenin I, Frémont V. Enhancing the Driver Awareness of Pedestrian using Augmented Reality Cues. 2016;p. 1298–1304.

[52] Wang Z. Human Vehicle Interaction in Automobile;.

[53] Foyle D. Attentional issues with superimposed symbology: formats for scene-linked. 2014;(March).

6.2 Appendix B: Experiment questionnaires

Pre-Test questionnaire

- What is your name? ……
- What is your age? ……
- What is your gender? m / f
- Do you have a driver's license? yes / no
- At which age did you obtain your first driver's license? ……
- On average, how often did you drive a vehicle in the last 12 months?

□ Every day	□ 4 to 6 days a week	□ 1 to 3 days a week	□ Once a month to once a week
□ Less than once a month	□ Never	□ I prefer not to respond	

- About how many kilometres (miles) did you drive in the last 12 months?

□ 0	□ 1 - 1,000	□ 1,001 - 5,000	□ 5,001 - 10,001
□ 10,001 - 15,000	□ 15,001- 20,000	□ 20,001 - 25,000	□ 25,001 - 35,000
□ 35,001 - 50,000	□ 50,001 - 100,000	□ More than 100,000	□ I prefer not to respond

— — — — — — — — — — — — — — —

How do you feel before the test?

Instructions. This questionnaire is concerned with your feelings and thoughts at the moment. Please answer every question, even if you find it difficult. Answer, as honestly as you can, what is true of you. Please do not choose a reply just because it seems like the 'right thing to say'. Your answers will be kept entirely confidential. Also, be sure to answer according to how you feel **at the moment**. Don't just put down how you usually feel. You should try and work quite quickly: there is no need to think very hard about the answers. The first answer you think of is usually the best.

Before you start, please provide some general information about yourself.

For each statement, circle an answer from 0 to 4, so as to indicate how accurately it describes your feelings **at the moment**.

Definitely false = 0, Somewhat false = 1,
Neither true nor false = 2, Somewhat true = 3, Definitely true = 4

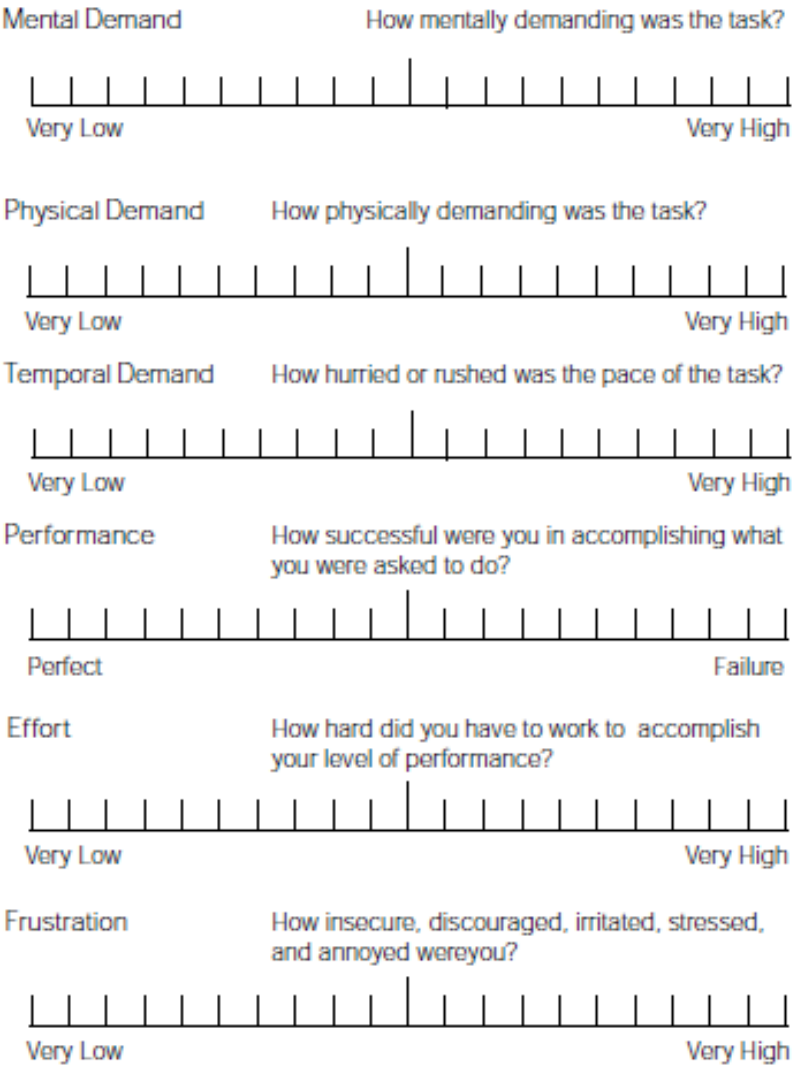
• The content of the task will be dull.	0	1	2	3	4
• I am determined to succeed on the task.	0	1	2	3	4
• My attention will be directed towards the task.	0	1	2	3	4
• I feel energetic.	0	1	2	3	4
• I will find it hard to keep my concentration on the task.	0	1	2	3	4
• I am motivated to try hard at the task.	0	1	2	3	4
• I feel tired.	0	1	2	3	4
• I feel bored.	0	1	2	3	4

How demanding was this task?

The evaluation you're about to perform is a technique that has been developed to assess the relative importance of six factors in determining how much workload you experienced.

You'll be presented with a series of scales. For each of the six scales, evaluate the task you performed by marking a vertical stripe on the scale: |

Each line has two endpoint descriptors that describe the scale. You can ask for clarification about the scales.



Note: this scale differs from the others.

What did you think of this augmented reality interface?

The goal of this interface is to help you safely and quickly respond when automation fails. Please indicate what you think of this interface.

•	Useful					Useless
•	Pleasant					Unpleasant
•	Bad					Good
•	Nice					Annoying
•	Effective					Superfluous
•	Irritating					Likeable
•	Assisting					Worthless
•	Undesirable					Desirable
•	Raising alertness					Sleep-inducing

How do you feel?

Instructions. This questionnaire is concerned with your feelings and thoughts while you were performing the task. Please answer every question, even if you find it difficult. Answer, as honestly as you can, what is true of you. Please do not choose a reply just because it seems like the 'right thing to say'. Your answers will be kept entirely confidential. Also, be sure to answer according to how you felt **while performing the task**. Don't just put down how you usually feel. You should try and work quite quickly: there is no need to think very hard about the answers. The first answer you think of is usually the best.

For each statement, circle an answer from 0 to 4, so as to indicate how accurately it describes your feelings **while performing the task**.

Definitely false = 0, Somewhat false = 1,
Neither true nor false = 2, Somewhat true = 3, Definitely true = 4

• The content of the task was dull.	0	1	2	3	4
• I was determined to succeed on the task.	0	1	2	3	4
• My attention was directed towards the task.	0	1	2	3	4
• I felt energetic.	0	1	2	3	4
• I found it hard to keep my concentration on the task.	0	1	2	3	4
• I was motivated to try hard at the task.	0	1	2	3	4
• I felt tired.	0	1	2	3	4
• I felt bored.	0	1	2	3	4

6.3 Appendix C: Consent form

Consent form for participants

Research Title: “Effect of naturalism in augmented reality based automation status interface for autonomous driving”

Researchers:

Jonas Pijnenburg – MSc student. Email: jtpijnenburg@gmail.com, +31 654318819

Christopher Cabrall – PhD student. Email: c.d.d.cabrall@tudelft.nl

Dr.ir. Joost C.F. de Winter – Supervisor. Email: j.c.f.dewinter@tudelft.nl

Location of the experiment:

Room 34 F-2-340
Faculty of Mechanical, Maritime and Materials Engineering Delft University of Technology
Mekelweg 2, 2628 CD Delft

Introduction: Please read this consent document carefully before you decide to participate. This document describes the purpose, procedures, and potential risks/discomforts. Your signature is required for participation. If you desire a copy of this consent form, you may request one.

Purpose of the study:

In future automated vehicles, the driver may be informed about the automation status by means of augmented reality on the windscreen. The purpose of this study is to analyse your opinion and responses to different types of augmented reality.

Duration: Your participation in this experiment will last approximately 40 minutes.

General procedures and instructions

Before the experiment starts: You will be asked to read instructions.

During the experiment: You will be watching video’s of driving situations. Your task will be to monitor lateral and longitudinal automation status and to respond appropriately to errors using the vehicle controls.

Paper-and-Pencil Tasks: At some points in time you will be asked to fill in a paper and pencil questionnaire regarding your feelings, thoughts, workload on the interface.

Risks and discomforts: There are no known risks for you in this study. Some people experience nauseousness in driving simulators. If at any point you begin

to feel uneasy for any reason, please do not hesitate to inform the experimenter as we would be happy to allow for a pause/break to counteract any such symptoms.

Confidentiality: All data collected in this study will be kept confidential and will be used for research and/or educational purposes only. You will not be personally identifiable in any future publications based on this work.

Right to refuse or withdraw: Your participation in this study is entirely voluntary. You have the right to refuse or withdraw from this experiment at any time.

Questions: For any questions, you can contact Jonas Pijnenburg (jtpijnenburg@gmail.com, +316 54318819). I have read and understood the information provided above. I give permission to process the data for the purposes described above. I voluntarily agree to participate in this study.

Name: .....
Signature: .....
Date:
..... / ..... / .....

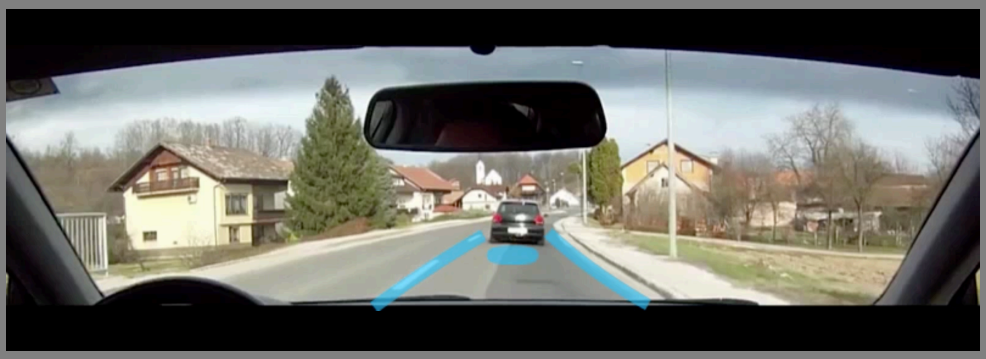
6.4 Appendix D: On-screen instructions

Welcome to this experiment on augmented reality and thank you for participating

Todays agenda:

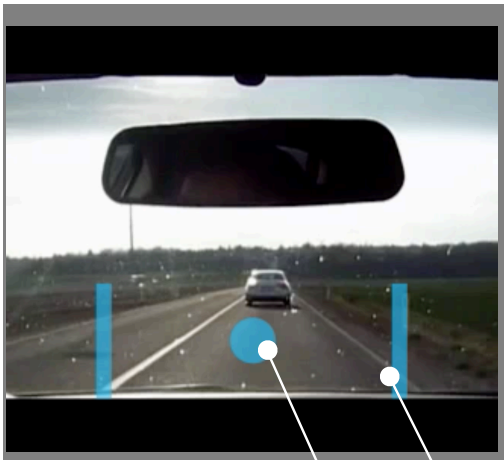
- Instructions
- Test 1 (7 minutes)
- Test 2 (20 minutes)

Press the top-left button to continue



- In this experiment, it will be simulated that you are in a self-driving car.
- You have to stay alert, to see if the car is able to drive well.
- Blue augmented interfaces will help you to do this.
- When these blue interfaces are on, all is well.

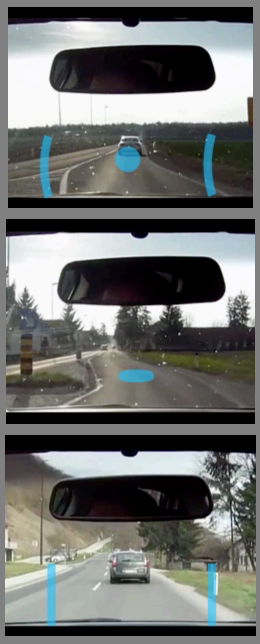
Press the top-left button to continue



The stripes are to show that the car knows its position compared to the lane markings.

The dot shows that the car knows its distance to objects in front of it.

Press the top-left button to continue



- **All is well:**
  The stripes and the dot are on.

- **Error:**
  The stripes are off. The car is not able to detect its lateral position.

- **Error:**
  The dot is off. The car is not able to detect its longitudinal position.

These errors occur due to various reasons, like sharp turns, a close object, uneven roads or road construction

Press the top-left button to continue

When an error occurs, your task is to:

- First grab the handles to activate manual steering
  (They are behind the steer)

1. Grab

- Then use the steer & pedals to safely respond to the situation

2. Steer & pedals

Please try to use the Grab handles to continue

When an error occurs, your task is to:

- First grab the handles to activate manual steering

1. Grab

- Then use the steer & pedals to safely respond to the situation

2. Steer & pedals

OK- the grab handles worked

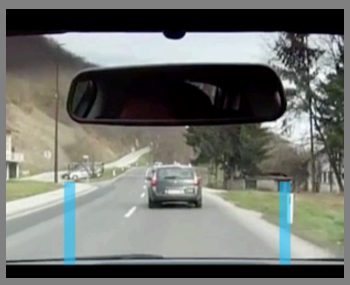
Press the top-left button to continue

To summarise, your task is to:

Do nothing

Stripes are off: error!

Respond!



Dot is off: error!

Respond!

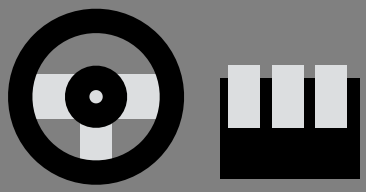
If anything about the task is unclear, don't hesitate to ask Jonas

Press the top-left button to continue

Test 1

- You will be shown multiple 5-second videos, with a black screen in between. These are not interactive.
- Some of these clips include automation status errors
- When you see an error, respond as quickly as possible:

1. Grab



2. Steer & pedals

Press the top-left button to start Test 1

Test 2 - 1

- You will be shown a 5 minute clip of a driving situation.
- At some point in time, automation errors will occur.
- When you see an error, respond as quickly as possible:

1. grab

2. Steer & pedals

Press the top-left button to start Test 2

Test 2 -2

- You will be shown a 5 minute clip of a driving situation.
- At some point in time, automation errors will occur.
- When you see an error, respond as quickly as possible:

1. grab

2. Steer & pedals

Press the top-left button to start Test 2

Test 2 -3

- You will be shown a 5 minute clip of a driving situation.
- At some point in time, automation errors will occur.
- When you see an error, respond as quickly as possible:

1. grab

2. Steer & pedals

Press the top-left button to start Test 2

6.5 Appendix E: Acceptance results

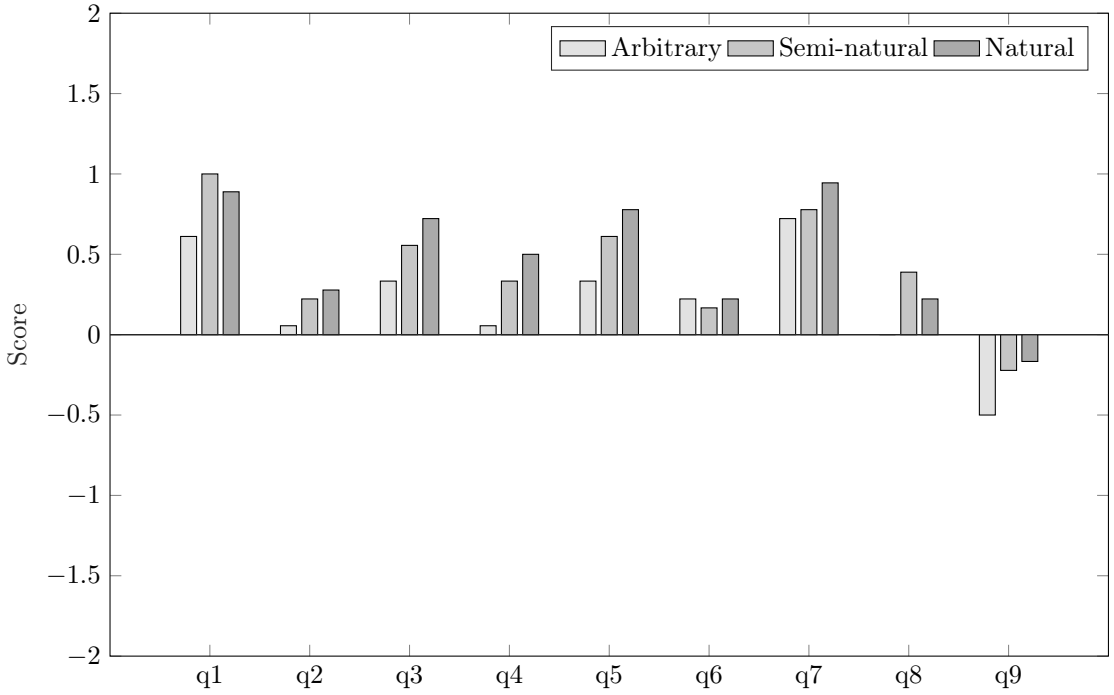


Figure 18: Full results for self reported Acceptance using the Van Der Laan Acceptance questionnaire. Breakdown of questions 1 through 9.

		F	p
q1	*Useful - Useless*	2.48	0.098
q2	*Pleasant - Unpleasant*	0.61	0.552
q3	*Good - Bad*	1.35	0.274
q4	*Nice - Annoying*	2.68	0.083
q5	*Effective - Superfluous*	2.84	0.072
q6	*Likeable - Irritating*	0.047	0.954
q7	*Assisting - Worthless*	0.64	0.535
q8	*Desirable - Undesirable*	1.52	0.233
q9	*Raising alertness - Sleep inducing*	1.21	0.312

6.6 Appendix F: Engagement results

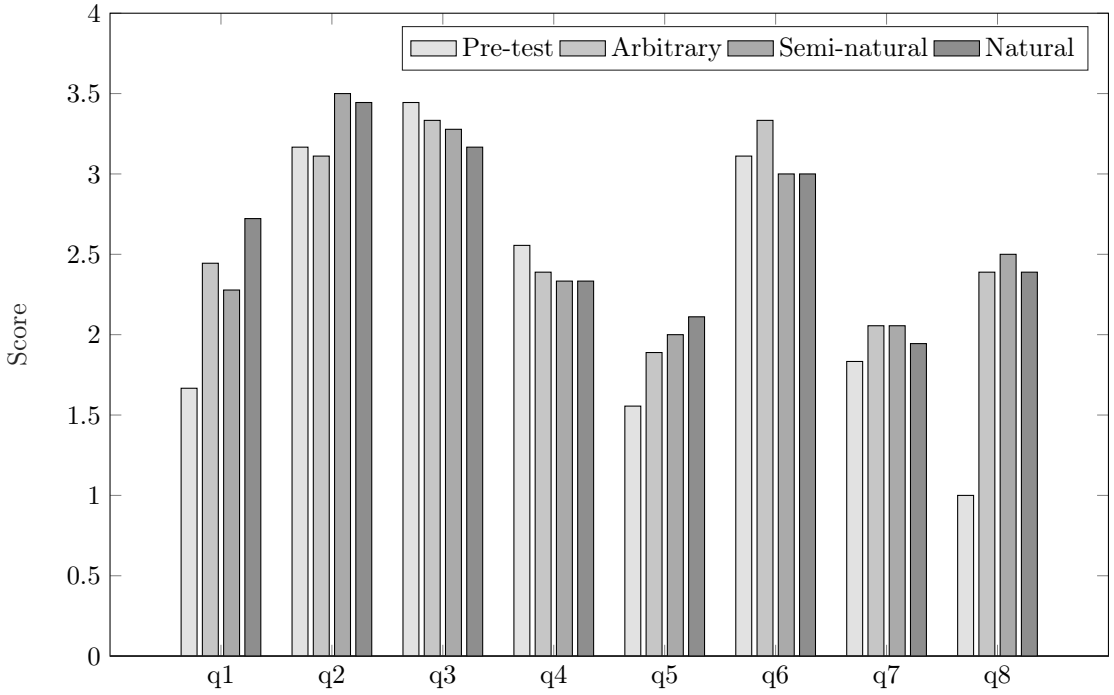


Figure 19: Full results for self reported Engagement using the engagement part of the DSSQ short questionnaire, breakdown of questions 1 through 8. For the Pre-test results, the questions were written in present tense. These were scored on a range from 0 - 4. Overall score is calculated by adding q2, q3, q4 and q6, subtracting q1, q5, q7 and q8 and adding 16, with a range of 0 - 32. A repeated measure one-way ANOVA was conducted, significant results are printed in bold.

		F	p
q1	*The content of the task was dull.*	**6.41**	**<0.001**
q2	*I was determined to succeed on the task.*	**4.30**	**0.009**
q3	*My attention was directed towards the task.*	0.95	0.424
q4	*I felt energetic.*	1.01	0.397
q5	*I found it hard to keep my concentration on the task.*	1.80	0.158
q6	*I was motivated to try hard at the task.*	1.92	0.137
q7	*I felt tired.*	0.71	0.552
q8	*I felt bored.*	**35.47**	**<0.001**

6.7 Appendix G: Supplementary material

Video & Document Repository on Google Drive

Name	Filename	Type
Exp. 1 video	exp1.mp4	Video
Exp. 2 videos [A1 - C3]	Exp2_[A1-C3].mp4	Video
Consent Form	consentform.pdf	Document
Instructions	instructions.pdf	Document
Questionnaires	questionnaires.pdf	Document
Repository link:	https://tinyurl.com/y7hwa3t5	

Statistical Analysis & Data repository on Google Sheets

Name	Type	File
Analysis	Contains statistical analysis of the data	https://tinyurl.com/ycvs4jqf
Objective measure	Contains and processes objective experiment data	https://tinyurl.com/y98yxs7z
Subjective measure	Contains and processes questionnaire data	https://tinyurl.com/y8aoth5a

TU Delft