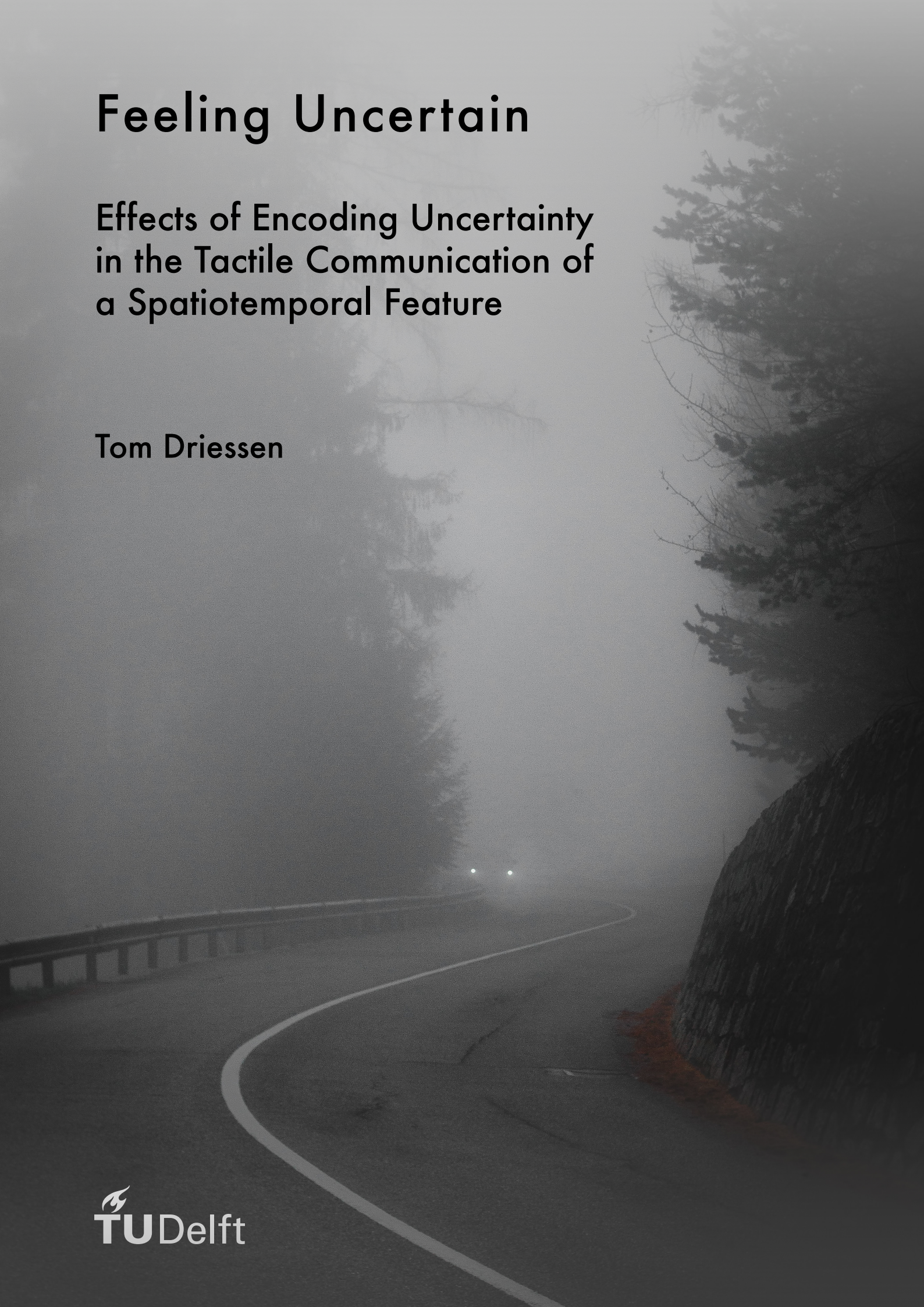


Feeling Uncertain

Effects of Encoding Uncertainty in the Tactile Communication of a Spatiotemporal Feature

Tom Driessen

TU Delft

Feeling Uncertain: Effects of Encoding Uncertainty in the Tactile Communication of a Spatiotemporal Feature

By

Tom Driessen

in partial fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

at the Delft University of Technology,
to be defended publicly on Tuesday August 27, 2019 at 15:00.

Supervisor:	Dr. ir. Joost De Winter, TU Delft
Thesis committee:	Dr. Pavlo Bazilinskyy, TU Delft
	Dr. Dimitra Dodou, TU Delft
	Matti Krüger MSc., Honda Research Institute EU

An electronic version of this thesis is available at http://repository.tudelft.nl/.
The digital appendix of this thesis is available at https://github.com/tomdries/feelinguncertain.

Preface

This thesis comes at the end of my eighth year of studying Mechanical Engineering in Delft. During this time I was equipped with analytical skills to break down complex systems into smaller, understandable parts and with design principles that shape creativity into new technologies. In my master program, my focus was drawn to one most intriguing systems of all: the human body. I chose the master track Biomechanical Design and learned about topics ranging from the mechanical properties of bones and tendons to the ways in which humans and machine interact with each other. This last research area is often named human-machine interaction and is situated at the intersection of engineering and psychology. As a passionate reader of books about the human mind, I decided to write my master thesis in this field.

I started an internship at the Honda Research Institute in Offenbach, Germany, and became part of an academic environment with research areas in human-machine cooperation, robotics and artificial intelligence. I learned how to conduct a human-subjects experiment, improved prototyping skills and operated in an international community of researchers and students. I experienced how technology can open up new ways to study human behavior, and how from the development of these technologies new inventions can come to be. The project resulted in a European patent application that will become public in October 2020 (application no. 19 168 465.3). On a personal level, the project has given me the confidence that I am able to conduct a scientific research and report about it in a correct and concise manner and it has made me realize the importance and fun of scientific work.

This work would not have been possible alone. I would like to thank Matti Krüger for his supervision during the internship and his continued support after. Matti, thank you always keeping your door open and being available for help at any time. I enjoyed our discussions that always resulted in new ideas and improvements for the experiment and the invention. Secondly I thank Joost de Winter for his thoughtful and critical feedback on the project and for being an excellent supervisor throughout the course of the entire project. I thank Christiane Wiebel, whose expertise in psychology and experience with human-subjects experiments were of great value and whose tips helped me keep an overview and give structure to the experiment.

I thank all the interns, PhD students and friends from Offenbach who have made my time in Germany one to never forget.

Finally, I thank my family for their endless support and encouragement.

Tom Driessen
Delft, August 2019

Contents

Nomenclature

Acronyms

FR	Front Right
FL	Front Left
BR	Back Right
BL	Back Left
HU	Human Uncertain
HC	Human Certain
MU	Machine Uncertain
MC	Machine Certain
MTTC	Minimum Time To Collision
TTC	Time To Collision
TTP	Time To Passing
uc	uncertainty communication

Chapter 1

Scientific Paper

Feeling Uncertain: Effects of Encoding Uncertainty in the Tactile Communication of a Spatiotemporal Feature

Tom Driessen
tomdrie@gmail.com
BioMechanical Engineering Department
Mechanical, Maritime and Materials Engineering
Delft University of Technology
Delft, Netherlands

Matti Krüger
matti.krueger@honda-ri.de
Honda Research Institute Europe
Offenbach am Main, Germany

Joost De Winter
j.c.f.dewinter@tudelft.nl
BioMechanical Engineering Department
Mechanical, Maritime and Materials Engineering
Delft University of Technology
Delft, Netherlands

Heiko Wersing
heiko.wersing@honda-ri.de
Honda Research Institute Europe
Offenbach am Main, Germany

ABSTRACT

An appropriate understanding of a machine's competences may be critical for safe use. Sharing measures of real-time function reliability could help users to adjust their reliance on machine capabilities. We designed a vibrotactile interface that communicates spatiotemporal information about surrounding events and further encodes a representation of spatial uncertainty. We evaluated this interface in a driving simulator experiment with varying levels of machine confidence linked to a simulated degradation of sensor signal quality and varying levels of human confidence induced through a degradation of visual feedback. A comparison between variants of the system indicated positive performance effects of providing uncertain information compared to a more conservative solution that only provided information above a specific confidence level. Subjective reports revealed a positive acceptance of uncertainty signaling in low-visibility conditions, comparable to acceptance ratings of a fully confident machine that accurately signaled the precise location of events.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**; **Haptic devices**; **User interface design**; **Mixed / augmented reality**.

KEYWORDS

Spatiotemporal Displays; Sensory Augmentation; Reliability Display; Uncertainty Encoding; Mobile HMI; Automotive HMI; Human-Machine Cooperation; Cooperative Driver Assistance;

1 INTRODUCTION

Modern cars can be equipped with sensory systems that provide the machine with perceptual capabilities surpassing human perception in various aspects. For example, camera systems provide the machine with continuous 360-degree vision, and Lidar provides sensory capabilities in the dark. Assistance systems can recruit these sensory capabilities to improve safety by providing the driver with supportive information during the driving task (e.g., lane departure warning, blind spot detection, navigation), or by letting the vehicle (partially) take over control (e.g., adaptive cruise control, automated lane keeping).

In dynamic conditions, the reliability of sensory systems may degrade due to changes in the environment. For example, the accuracy of Lidar measurements tends to decrease in the rain [8] and car manufacturers warn about a reduced reliability of sensors in tunnels (e.g. [1], p. 96). Since the driver cannot be expected to have an understanding of the underlying mechanisms (or the mere existence) of these sensory systems, users may benefit from the availability of measures of machine reliability. A machine could assess these measures of uncertainty by itself; the level of uncertainty may be based on signal spread or disagreement between different sensory systems. A machine that can share such measures of self-assessed uncertainty could help users to dynamically adjust their level of trust in the automation to appropriate levels [19] and in consequence use this information to appropriately distribute attention and available resources.

1.1 Background

Attempts to communicate machine uncertainty have mainly focused on visual displays [2, 5, 9, 18, 20]. A disadvantage of using the visual modality is that it is not continuously available for input: communication depends on the gaze direction of the user and the visual modality is being competed for by other activities such as monitoring the environment. Modalities that are less subject to these problems are the haptic, auditory, and olfactory modalities. Recent studies have investigated the use of vibrations [17] as well as smell [23] to share measures of system uncertainty with the driver.

Vibrotactile displays have been a topic of research in the context of spatial perception assistance. General guidelines to the designs of vibrotactile displays have been established by Van Erp [22].

The current study uses a vibrotactile display that has the purpose of enhancing the user's spatial awareness by supplying directional

stimuli associated with the location of surrounding entities. Comparable applications found in literature have based the stimulus intensity on distance [3, 4] or spatiotemporal measures [15]. We have found no applications that have encoded measures of uncertainty into the signal.

Presenting system uncertainty fits in the domain of cooperative automation frameworks, which challenge designers to regard assistance functions or autonomous features as cooperative partners or team agents, rather than as tools, e.g. [6, 10, 14, 16]. Among ten challenges to make automation a team player, Klein et al. [14] listed the team agent's ability to "make pertinent aspects of their status and intentions obvious to their teammates" (p. 93). Also, they mentioned the ability to "direct each other's attention to the most important signals, activities and changes" (p. 94).

1.2 Current Study

Here we aim to support a driver in developing an understanding of the environment through sensory augmentation. In particular, we map machine observations about safety-critical spatiotemporal features to human-perceivable tactile stimuli. Additionally, we encode a measure of machine uncertainty in such stimuli to let the user obtain a better understanding of the accuracy of the suggestions by the machine.

The sensory enhancement device is currently adopted in a simulated automotive environment, which serves to study behavioural influences of the assistance system. In particular, we study the system's influence on safety, attention distribution and measures of user acceptance, understanding and workload.

This study may provide new insights about the design of uncertainty displays, sensory augmentation using vibrotactile feedback, transparency about spatial uncertainty and the concept of human-machine cooperation.

2 METHOD

2.1 Hardware

The experiment was conducted in a static driving simulator (Figure 1) with controls for steering, braking and accelerating. Gear-shifting/transmission was set to automatic mode. Three display panels (50 inch diagonal, 1080p each) presented the driving scenario and the remaining parts of the interior (dashboard, instrument cluster, mirrors) at 60 Hz, using the SILAB 5.1 driving simulation software developed by WIVW GmbH (Würzburg Institute for Traffic Sciences, Germany). Participants were equipped with a wearable 120 Hz monocular eye-tracker from Pupil Labs [13] and a waist belt containing 16 equally spaced vibromotors (Feelspace GmbH). The firmware was customized for the purpose of the experiment.

2.2 Communication Signals

2.2.1 Signal For a Confident Machine. The tactile communication was implemented with a signaling mode similar to the interface used in the experiments by Krüger et al. [15]. It signaled to the driver from which lane(s) vehicles were approaching by activating predefined vibromotors that were corresponding to the direction of the lane (Figure 2). The vibration intensity of the signal was a function of the time to collision (TTC) between the object (surrounding vehicle) and the driver's vehicle. TTC is a spatiotemporal measure

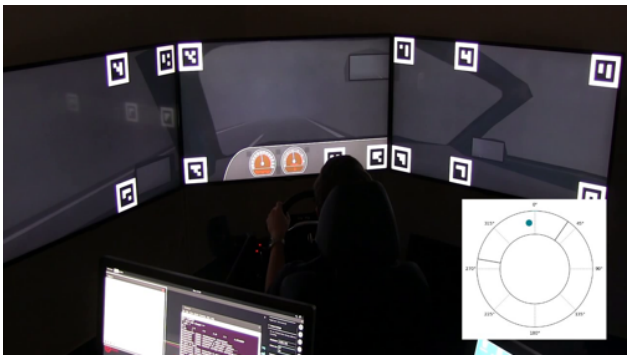


Figure 1: Driving simulator setup in foggy tunnel scenario, with uncertainty signaling visualized for the experimenter.

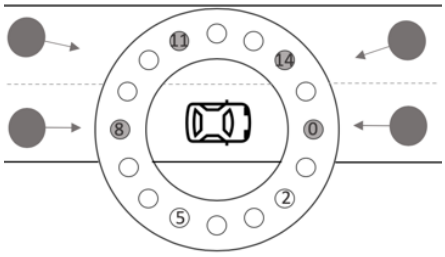


Figure 2: Schematic of the belt in an example situation where from every lane direction an object (large gray dot) is approaching with a TTC value under 9 s. Vibromotors nr. 0, 14, 8 and 11 (grey small dots) would activate in this case. If the ego-vehicle drove on the left lane, the activations would occur at vibromotor 0, 2, 5 and 8. Note that the vibromotors on the rear were spaced further apart, as the experimenters felt that the signals from the rear were easier to distinguish with a larger spacing between the actuators.

that is a function of the distance and relative velocity between entities. It is defined as the time it would take until a collision occurs if two entities would not adjust their current velocities and direction of traveling. As lower TTC indicates more dangerous situations, the stimulus intensity was based on the inverse of TTC.

The onset of vibrations occurred when the TTC dropped below a threshold of 9 seconds, at a vibration intensity that was identified by the experimenters as the lowest perceivable intensity. The vibration intensity increased as the TTC dropped, so that a lower TTC implied a higher stimulus saliency. When the TTC was zero (a collision), a stimulus reached its maximum intensity, which was equal to the maximum stimulus intensity of the Feelspace belt.

The currently explained communication mode can give appropriate signals about the location of approaching objects as long as the machine had precise knowledge about the location and velocity of these objects. The conveyed message that is being signaled is analogous to pointing to something and saying: 'something is exactly over here'. We refer to this signal as *the precise signal.*

2.2.2 Signal For an Uncertain Machine. In addition, the machine was equipped with a communication mode capable of informing the driver about the presence of approaching[1] objects when information is not conclusive enough to state the current lane of the approaching object, but where information is reliable enough to state that 'an object is somewhere within a range, though there is not enough evidence to say where exactly'. We refer to this signal as *the uncertainty signal.*

Figure 3A shows a schematic of the uncertainty signal, with one vehicle approaching from the front. The exact location of this vehicle is unknown, as illustrated by the strip of overlapping grey dots. The machine does have information that a vehicle is approaching, but it is uncertain on which lane it is located. To communicate this uncertainty, a dynamic vibration pattern is activated, successively activating neighboring vibromotors in the clock- or counter-clockwise direction. The initial vibrator position and direction is chosen randomly from the available vibromotors within this uncertainty range. The dynamic pattern travels over a pre-defined range that represents the overlap between the two lanes between which is doubt. The pointer oscillates from between the two borders with a constant frequency (1.0 Hz, from start-to-start point). The next vibromotor activates at the same instance that its predecessor switches off (Figure 3B). The pointer continues to bounce between these borders until either one of two events occurrs: (1) the TTC becomes greater than a specific threshold (in this case nine seconds), in which case the signal disappears, or (2) a reliable estimate of the current lane of the approaching vehicle becomes available. In the latter case, the width of the range converges to zero, to the static location associated with the object's lane as in Figure 2. The dependence of vibration intensity on TTC was the same as it was for the precise signal.

For a visualization of the uncertainty signal, we refer to the supplementary video (See Appendix A for details).

2.3 Driving Scenarios

Five simulator scenarios were designed. Two scenarios had the purpose of familiarizing the participant with the driving simulator and the tactile interface and the other three scenarios were designed with the aim of independently modulating the uncertainty of the vehicle's observations and the uncertainty of the human's observations.

Human uncertainty was induced by substantially limiting visibility in the driving scene through fog, making the look-ahead distance about 33 m. The machine was set to be uncertain during rain or when driving through a tunnel. In these scenarios machine uncertainty was simulated by making the current driving lane of an object unknown to the machine when an object was more than 33 meters away from the ego-vehicle. The only information that remained available to the machine was the longitudinal distance to the object and whether the object was in front or behind the vehicle. Participants were orally informed by the experimenter about these machine limitations before the start of each driving scenario, as follows: "In this section you will drive through rain/a tunnel.

[1]We regard approaching in the sense of reducing the relative distance in any direction regardless of who is faster (being approached by someone = approaching someone).

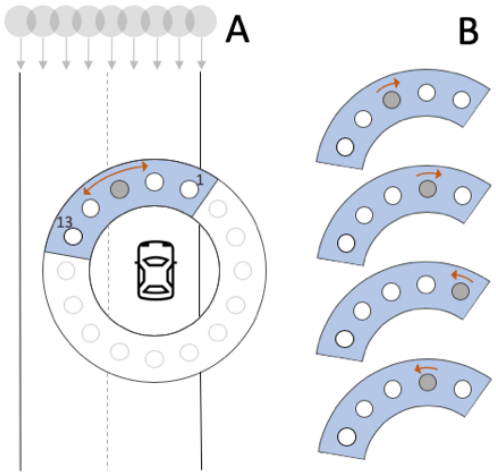


Figure 3: Uncertainty signal for an object approaching from the front on a two-lane road (A). The stimulus traveled between the borders and bounced back in the other direction as it hit one of the borders (B). The width of the range was chosen to be between the vibromotors that were allocated for the static signal (Figure 2) plus one extra vibromotor on each side. Thus, in the example in this image, the signal bounced between vibromotors 13 and 1.

Therefore, the vehicle is less certain about the locations of vehicles that are further away".

The instructions to the participant were to maintain a speed of 120 km/h where possible and avoid collisions with other vehicles. All scenarios consisted of a straight two-lane highway. The following subsections further detail the design of the scenarios.

2.3.1 Familiarization Scenarios. Before conducting the experiments, participants encountered two scenarios to accommodate them with the simulator and the vibrotactile belt functionality.

The first familiarization procedure was carried out according to guidelines specified by Hoffmann and Buld [11]. This procedure was aimed at reducing the probability of causing simulator sickness by gradually increasing exposure to virtual accelerations. Longitudinal accelerations were introduced by asking the participant to accelerate to 30 km/h and then come to a full stop. This was repeated, increasing the speed to 50, 80 and 120 km/h. Finally, participants were asked to swerve between cones placed alternating between the right and left lane, introducing lateral accelerations at a speed of approximately 30 km/h. The total duration of this scenario was approximately four minutes.

The second familiarization scenario familiarized the driver with the spatiotemporal proximity signaling functionality of the belt in a scenario where the machine was certain (precise signal). The scenario consisted of a two-lane highway on a sunny day. Participants were not informed about the functionality of the belt and were asked to maintain a speed of 120 km/h where possible. Since vehicles on the passing lane were designed to drive faster than the target speed, the task was most easily fulfilled by driving on

Figure 4: Visibility in the Foggy scenarios. Vehicles disappear into the fog at a distance of approximately 33m.

the rightmost lane. However, vehicles on the right lane that were trailed by the ego-vehicle would occasionally slow down, forcing the participant to either overtake via the left lane or brake to avoid a collision. These instances ensured that the time to collision between the ego-vehicle and its surroundings dropped below the threshold value of 9.0 seconds, invoking exposure to the signal. After five minutes of driving, participants were asked to park their car on the emergency lane and the belt exploration scenario was stopped. Participants were then asked what they thought the tactile stimuli communicated, and they were informed about the true nature of the assistance function. This scenario was similar to an experimental scenario by Krüger et al. [15], who found that participants were able to develop an intuitive understanding of the stimuli within four minutes of system exposure. Similar rapid user understanding times for directional tactile displays were described by Cassinelli et al. [3] and Hogema et al. [12].

2.3.2 Foggy Road: Machine Certain, Human Uncertain (MC-HU). The foggy road scenario was simulated as a night-time scenario, designed to make the human uncertain by inserting a dense fog field and disabling lights of surrounding traffic. This limited the look-ahead distance to about 33 m (Figure 4), corresponding to a lookahead time of about 1.0 s assuming the driver drove at approximately the target speed.

Machine observations were not affected by the mist or darkness, so a precise signal was communicated for vehicles driving at any distance away from the ego-vehicle.

Approaches of surrounding vehicles could be triggered by the experimenter. When a command was given, a vehicle started approaching from behind the fog barrier from one of the four possible lane directions, driving at a speed of 160 km/h (for vehicles from the rear) or 80 km/h (vehicles from the front). Eventually the vehicle would overtake (or be overtaken by) the ego-vehicle, assuming the ego-vehicle drove around the target speed of 120 km/h. After a vehicle passed and disappeared into the fog again, and the experimenter confirmed that the participant had reached a speed of 120 km/h again, the next vehicle was launched. This was repeated until 14 vehicles were launched.

Cars that approached from the rear on the right lane would change lanes and overtake the ego-vehicle at a distance of 30m.

The direction from which cars approached was a randomized sequence, re-used in each experiment for comparability. The order of the encounter directions is listed in Table 1.

2.3.3 Foggy Tunnel: Machine and Human Uncertain (MU-HU). The foggy tunnel scenario was identical to the foggy road scenario, except for the addition of a tunnel that ran for the entire course.

Table 1: Order of launched vehicles. Vehicles annotated with an * approached at 0 km/h instead of 80 km/h. Due to comparability issues they were excluded from the analysis.

1	2	3	4	5	6	7	8	9	10	11	12	13	14
FL	FL	FL	BL	BR	FR*	FL	BL	FR*	BR	BL	BR	FR	FR

The driver visibility was again approximately 1.0 s. The tunnel had the effect that the machine became uncertain about the locations of vehicles that were outside of the visible region (>33 m away).

The traffic definitions were identical to the Foggy Road scenario, for comparability between foggy scenarios.

2.3.4 Rain: Machine Uncertain, Human Certain (MU-HC). The rain scenario consisted of a straight road on a rainy day. The rain was visually present, though at an intensity at which it was assumed not to be of influence on the driver's visual perception. The reliability of the machine was negatively affected by the rain, in the same manner as it was in the foggy tunnel scenario.

In the rainy scenario, the driver's field of view was not obstructed by fog. The traffic setup from these scenarios could not be re-used since there was no fog to mask the fact that cars suddenly appeared on the road ahead. A different traffic profile was designed for the rain scenario, which is explained in Figure 5.

2.4 Participants

Fourteen drivers (1 female) between 21 and 41 years old (M = 29.1, SD = 5.4) participated this study. All participants reported that they had (corrected-to) normal vision and held a valid driving license for an average of 11 years. Two participants reported that, over the past year, they had driven less than once a month. Two drove once a month to once a week, three drove one to three days a week, and seven drove four days a week or more. For nine participants driving a car was their primary mode of transportation. Three participants reported mainly using motorcycles and the remaining participants relied primarily on walking, public transportation or other methods. All participants were students or employees active at XXXblank for blind reviewXXX.

2.5 Experimental Design

In two scenarios (Foggy Tunnel and Rain), the machine became uncertain about its observations. These scenarios were encountered twice by each participant: once without and once with uncertainty communication functionality enabled. When uncertainty communication was disabled, the machine only communicated the precise

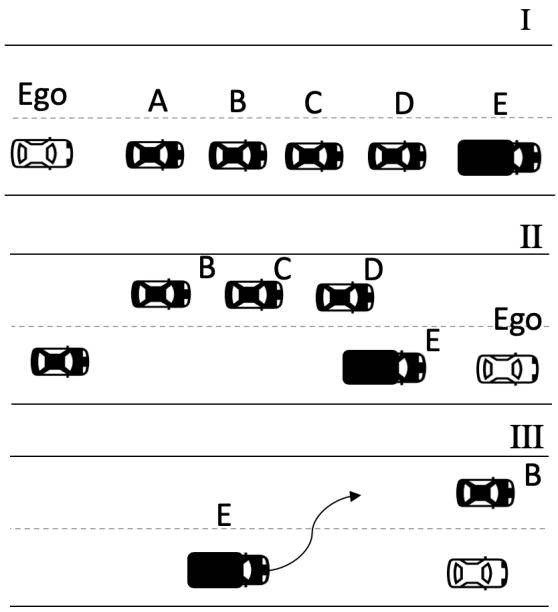


Figure 5: Traffic definition in the rain (MU-HC) scenarios. Five vehicles were driving on the right lane at 80 km/h, spaced 250 m apart (I). The ego-vehicle could stick to the target speed (120 km/h) by overtaking the vehicles. When the front truck (E) was overtaken, a trigger point was activated that made the trailing cars B, C and D switch to the left lane, and adjust their speed to 160 km/h (II). This resulted in B, C and D eventually overtaking Ego from the rear. When D passed the ego-vehicle (III), the leading truck (A) accelerated to 160 km/h, and it changed to the left lane if it came within a distance of 30 meters of the ego-vehicle.

signal as soon as a reliable estimate of an object's position became available.

In the Foggy Road scenario, the machine had an accurate estimate of the position of vehicles at any distance away from it, thus it could always communicate the precise signal. Since there was no machine uncertainty in this condition, there was no need to conduct this scenario twice. This resulted in a total amount of five experimental conditions, whose characteristics are summarized in Table 2.

To rule out potential learning effects, the order in which these experimental conditions were conducted was varied between participants. Half of the participants started with the two uncertainty communication conditions and half without. Also, we ensured that foggy scenarios and rain scenarios were alternated, as test runs indicated that the foggy scenarios were most intense in terms of mental demand. The amount of time consecutively driven with high mental demand was minimized by this ordering.

An indicative time schematic of an experimental session is given in Table 3.

Table 2: Experimental conditions. In the machine visibility column, '33m + uc' means that the machine observes precise positions up to 33m. Beyond that, it communicates the uncertainty signal.

condition	scenario	uncertainty signaling	machine visibility	human visibility
MC-HU	foggy road	not necessary	not limited	33 m
MU-HU	foggy tunnel	not available	33 m	33 m
MU-HC	rain	not available	33 m	not limited
MU-HU-uc	foggy tunnel	available	33 m + uc	33 m
MU-HC-uc	rain	available	33 m + uc	not limited

Table 3: Approximate time scheme.

Informed consent and demographics questions	10 min
Simulator + belt familiarization scenarios	10 min
Eye tracker installation + calibration	5 min
Experimental conditions (5x) + questionnaires	30 min
Room for open comments by participant	5 min
Total	60 min

2.6 Dependent Variables

The following variables were recorded:

Trial safety: MTTC. We consider the time fragments leading up to an overtake of (or an overtake by) a surrounding vehicle as individual *trials*. (see section 2.7.1 for details). For each trial, the minimum time to contact (MTTC) was recorded as an indication of trial safety. MTTC indicates safety, as higher TTC values may indicate (1) that the ego-vehicle adjusted its speed to match that of an approaching vehicle or (2) that the ego- vehicle changed lanes to an unoccupied lane, causing the TTC to increase towards infinity.

Trial safety: Steer angle and brake depression. On the highway, high steering angles and high brake depression can be indicators of unsafe situations. They can also be used to identify if and when participants responded to a signal and whether they chose to brake and/or overtake.

Gaze distribution. Eye-tracking data was collected to evaluate if the onset of the uncertainty signal caused shifts in visual attention towards the direction of the presented signal. We defined the dependent measure *front gaze ratio* as the ratio of the amount of gaze points in the front window versus the total amount of gaze points in the mirrors and wind shield (Equation 1). A higher front gaze ratio indicates the user allocated more attention towards the front, a lower front gaze ratio indicates more attention towards the rear. See Section 2.7.3 for details about the analysis of the eye tracking data.

$$\text{front gaze ratio} = \frac{\text{gaze count on windshield}}{\text{gaze count on windshield + mirrors}} \quad (1)$$

Subjective Measures. After each experimental condition the NASA Task Load Index (Raw-TLX, [7]) assessment was conducted as well as the Van Der Laan usefulness/satisfaction questionnaire [21]. Furthermore, after every experimental run the participant was asked to rate the following items on a 5-point Likert scale (strongly disagree to strongly agree). These statements allowed us to check if the environmental modulations indeed were of influence on the human's confidence, if the participants had understood that the machine became less certain and if they found that the machine communicated this. These questions may further serve as diagnostic indicators in case of unexpected participant behaviour.

(1) The other road users made me unconfident
(2) The weather conditions made me unconfident
(3) The signals from the belt made me unconfident
(4) The machine was sometimes uncertain about the exact location of a vehicle
(5) The machine told me it was uncertain about the exact location of a vehicle
(6) I relied on what I perceived with my eyes
(7) I relied on what I perceived through the belt
(8) I had trust in my own capabilities

2.7 Analysis

2.7.1 Extracting trials from the recordings. The recordings were split up into smaller time fragments that we will call *trials*. A trial was created for every vehicle that overtook or was overtaken by the ego-vehicle. The end point of a trial was defined as the exact moment a vehicle passed the ego-vehicle. The starting point of a trial was defined as the moment where time to passing (TTP) of a surrounding vehicle dropped below 9 seconds. Here, TTP is a dynamic variable that was defined as the time it would take until two vehicles would pass each other if they would maintain their current speeds. TTP is similar to time to contact (TTC), however TTP is also defined when two vehicles are not on a collision trajectory. This means that two vehicles do not need to drive on the same lane in order for TTP to take on a value. This was a helpful property that allowed us to define the starting point of trials where the ego-vehicle and the surrounding vehicle were not initially driving on the same lane.

2.7.2 Analysis of Minimum Time to Collision (MTTC). For the analysis of trial safety measures, we focused on the conditions in which the visible field of the human was limited (foggy). In particular, we regarded the encounters with vehicles that approached from the front right lane that were driving with 80 km/h. Since the driver was asked to drive on the right lane when possible, these were the vehicles that required action by the driver in order to avoid a collision (change lane or brake). For each trial, the MTTC was calculated. Since there were three of these trials per condition, we took the average of the three obtained MTTC values for the analysis.

For the MU-HU (Foggy Tunnel) conditions, comparisons were made between the uncertainty-communication-disabled condition (MU-HU) and the uncertainty-communication-enabled condition (MU-HU-uc). It was hypothesized that MTTC scores would be higher for the condition where uncertainty communication was enabled, since the machine was capable of notifying the participants earlier that a slow vehicle was driving ahead.

Another comparison was made between the foggy tunnel with uncertainty communication enabled (MU-HU-uc) and the foggy road condition (MC-HU). In the latter condition, the machine was always confident about the exact lane of an object and thus communicated a precise description of the position of objects to the user. It was hypothesized that the driver would use the precise signals more appropriately then the uncertain signals, as the driver gets information from which he/she could infer the lane of the object. We hypothesized that this would be apparent from higher MTTC scores for the MC-HU condition than for the MU-HU-uc condition.

2.7.3 Analysis of Eye Tracking Data. We defined areas of interest around the front window and the three mirrors and recorded in which of these areas the estimated gaze point of the participant was. We were interested to see if the presence of the uncertainty signal caused the driver to shift visual attention towards the direction that was being signaled by the assistance. The measure that we used to indicate a shift of attention towards or away from the front is the front gaze ratio (Equation 1). We calculated the front gaze ratio for trials where a vehicle approached from the rear. Each experimental run contained three of these trials. Per participant, the mean score of these trials was used for statistical comparisons. We expected that in conditions with uncertainty communication, the drivers' gaze would be focused more on the mirrors (lower front gaze ratio) then for conditions without uncertainty communication.

3 RESULTS

Figure 6 shows for each trial in human-uncertain conditions the development of TTC, speed, steer angle, brake depression and driving lane. These plots indicate earlier responses of participants in situations where there was early communication, either in the form of uncertainty signaling (middle column) or precise signaling (right column). More aggressive braking for the late precise signal (left column) can be observed, likely because participants were informed later about oncoming traffic. They had less time to avoid the obstacle. These figures indicate that participants understood that the uncertainty signal meant that something was approaching, and that participants were able to use the supplied information for the benefit of safety (in terms of TTC).

3.1 MTTC

Figure 7 displays the MTTC scores for the conditions in which the human was uncertain. A paired t-test was used to compare the uncertainty communication in human-uncertain scenarios (MU-HU-uc) with the baseline signal (late, certain communication; MU-HU) and with the all-knowing variant (MC-HU). On average, the MTTC was higher for the MC- HU condition (M = 3.9 s, SD = 1.11) than for the MU-HU-uc system (M = 2.59 s, SD = 0.88); t(13) = 4.73, $p < .001$.

Compared to the system without uc (M = 1.24s, SD = 0.46), the uncertainty communicating system had a higher MTTC on average; t(13) = 4.36, $p < .001$.

These results indicate that, in terms of MTTC values, the safest condition was the condition with a certain machine. It is likely that the participant correctly understood the precise signal and changed lanes or adjusted its speed the earliest. However, in conditions with less optimal sensory conditions, the uncertainty communication

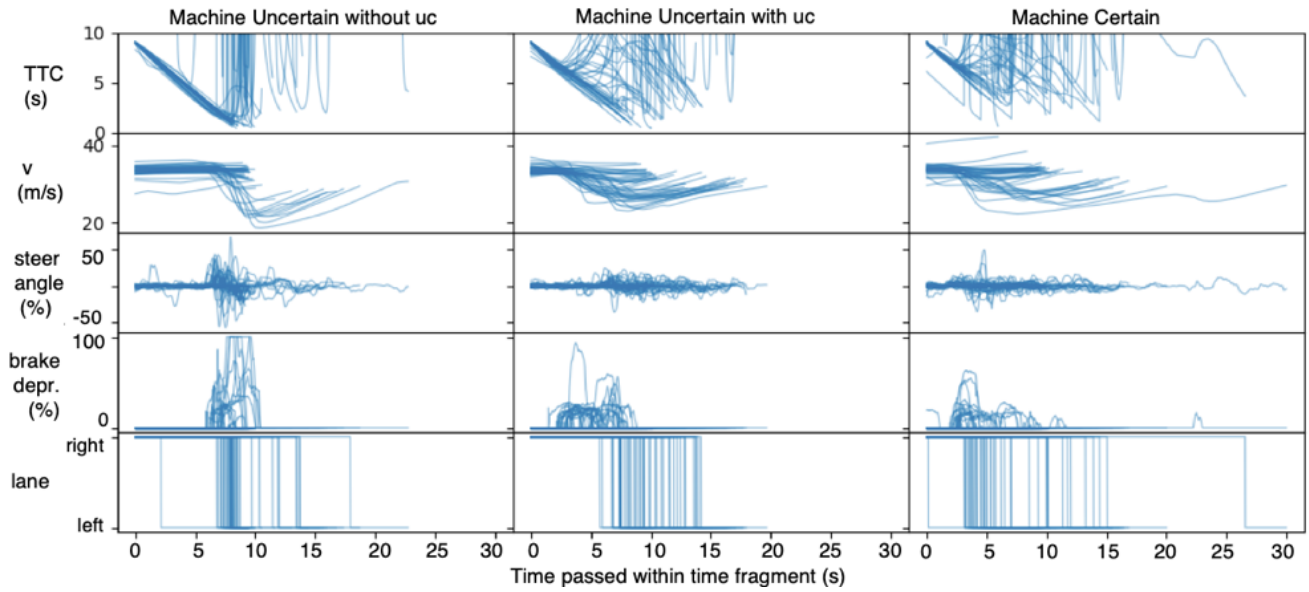


Figure 6: Raw data plots for trials in conditions with an uncertain human.

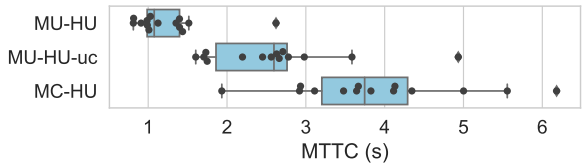


Figure 7: MTTC scores for human-uncertain conditions (n = 14).

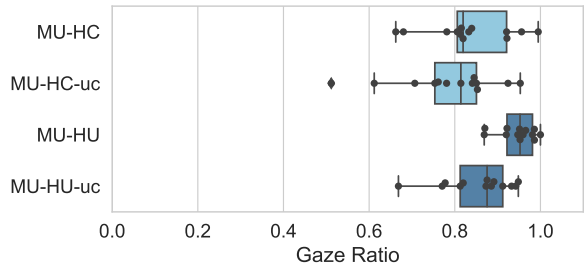


Figure 8: Gaze ratio for conditions in which the machine was uncertain and for trials in which vehicles were approaching from the rear. Lower values indicate more gazing towards the mirrors. Due to failed eye tracking recordings, n = 13 (instead of 14) for all conditions.

signal is superior to a machine only capable of signaling specific, reliable observations.

3.2 Eye Tracking

Figure 8 shows the ratio of gaze points on the front (front window) divided by front+back (front window + mirrors).

For the human-certain conditions (rain), a one-sided Wilcoxon signed rank test did not reveal that uncertainty communication from the rear significantly lowered the front gaze ratio: (MU-HC vs. MU-HC-uc: p = .066, see light blue boxplots). In the human-uncertain conditions, a significant decrease in front gaze ratio was shown (MU-HU vs. MU-HU-uc: p < .001, see dark blue boxplots). These results indicate that the signal successfully directed visual attention towards the direction that the signal appeared from in conditions where the human was uncertain.

3.3 Subjective Reports

Usefulness and satisfaction ratings were conducted for the three conditions where the machine had the ability to communicate about vehicles that were more than 33 m ahead. In the MU conditions, this was in the form of the uncertainty signaling, in the MC condition, this was in the form of a precise signal. Figure 9 summarizes the responses. A Friedman test for the usefulness and satisfaction indicated a significant difference between the three conditions, ($\chi^2(2)$ = 21.93, p < .001 for usefulness, $\chi^2(2)$ = 11.41, p = .003 for satisfaction). Post-hoc Wilcoxon signed rank tests showed that usefulness was rated significantly higher for the conditions where the human visibility was limited by fog (MC-HU and MU-HU) than for the rainy condition, in which only the machine was uncertain (MC-HU vs. MU-HU: p = .289, MU-HU vs. MU-HC: p < .001, MC-HU vs. MU-HC:

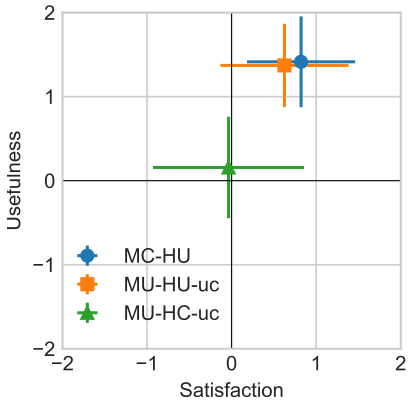


Figure 9: Mean usefulness and satisfaction scores of the assistance functionality in MC-HU (Foggy Road), MU-HU-uc (Foggy Tunnel), MU-HC-uc (Rain). Error bars display the standard deviation. The mean (SD) usefulness scores in these conditions were 1.41 (0.54), 1.37 (0.50), 0.16 (0.50), respectively. For the satisfaction scores: 0.82 (0.64), 0.62 (0.76), -0.04 (0.76), respectively.

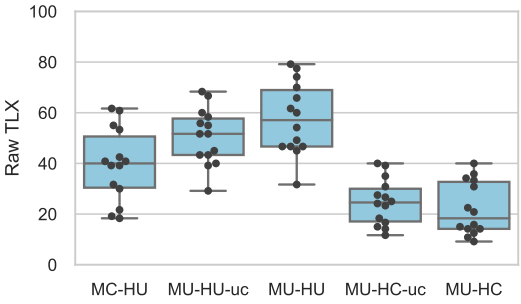


Figure 10: NASA Raw TLX scores per condition. Scores of each sub-question were averaged to obtain the overall RTLX score, ranging from 0-100.

$p < .001$). This was also observed in the satisfaction ratings (MC-HU vs. MU-HU: $p = .219$, MU-HU vs. MU-HC: $p = .004$, MC-HU vs. MU-HC: $p = .001$).

We observed higher NASA TLX workload ratings in HU conditions than in HC conditions (Figure 10).

Secondly, a Wilcoxon signed rank test showed a significantly lower workload rating for the MC-HU condition when compared to the MU-HU-uc condition ($p = 0.032$). This further indicates that, in human uncertain conditions, participants had less difficulties when receiving long-range specific signaling then when receiving uncertainty signaling.

Table 4: Mean (SD) answers to questionnaire, rated on a 5-point Likert scale. Scores range from 0 (strongly disagree) to 4 (strongly agree). Items: 1. other road users made me unconfident, 2. weather conditions made me unconfident, 3. signals from the belt made me unconfident, 4. The machine was sometimes uncertain about the exact location of a vehicle, 5. The machine told me it was uncertain about the exact location of a vehicle, 6. I relied on what I perceived with my eyes, 7. I relied on what I perceived through the belt, 8. I had trust in my own capabilities.

	MC-HU	MU-HU-uc	MU-HU	MU-HC-uc	MU-HC
1	1.86 (1.15)	2.00 (1.00)	2.50 (1.12)	0.43 (0.49)	0.64 (0.61)
2	3.00 (1.20)	3.36 (0.81)	3.64 (0.48)	1.00 (0.53)	1.00 (0.38)
3	0.57 (0.62)	1.07 (0.80)	1.93 (1.03)	1.36 (1.11)	1.10 (0.30)
4	0.86 (0.99)	3.29 (0.70)	2.77 (1.31)	3.43 (0.62)	1.80 (1.40)
5	0.80 (0.75)	3.17 (0.90)	0.67 (0.75)	3.40 (0.80)	1.00 (0.71)
6	2.36 (1.04)	2.00 (1.00)	3.50 (0.63)	3.79 (0.41)	3.50 (1.05)
7	3.64 (0.48)	3.36 (0.61)	1.36 (1.04)	2.14 (0.91)	1.18 (1.11)
8	2.50 (1.05)	2.36 (0.61)	2.29 (1.03)	3.57 (0.49)	3.36 (1.04)

Responses to the eight Likert items are listed in Table 4. They indicate that the manipulations worked as we indented; participants reported that they were less confident in the foggy scenarios and they seemed to have understood when the machine was uncertain. At the end of the experiment, participants were asked what they thought about the system and if they had any comments they wished to write down. Two participants commented that they found the task unrealistic since they would never drive 120 in the real world in this type of fog. Two participants commented that they wanted different information from the back and front, of which one further specified that this was since (s)he did not rate cars from behind as dangerous. One participant noted that (s)he would like to have such a thing in his/her car, added with the comment that (s)he would 'probably be tempted to drive faster'. One participant commented that (s)he 'could never tell if the device was intentionally conveying uncertainty or if its behavior when it was uncertain simply made it impossible to infer'. One other participant noted: 'The uncertainty expression of the system was very intuitive! It was clear that the system was uncertain about the exact location of other cars.'

4 DISCUSSION

In the current study, we investigated the effects of communicating spatial machine uncertainty via a vibrotactile pattern. We evaluated the influence of the uncertainty communication on safety, attention allocation and subjective measures. To do so, we varied the competences of the machine by simulating sub-optimal sensory conditions for the machine (tunnel, rain). In addition, we manipulated human confidence by inserting dense fog fields. A prerequisite to this study was that these environmental manipulations indeed had the effect that we intended. Subjective reports suggest that successful manipulations took place. Participants perceived higher workload in all the foggy conditions and further subjective reports

indicated that the fog made participants uncertain (Table 4, q2) and that they had a diminished self-reliance (Table 4, q8) .

Implications for safety. Changes in driving behaviour for the benefit of safety were observable; the uncertainty signaling induced earlier lane-changes or braking responses in overtaking situations in comparison to situations where no signaling was present (Figure 6). This observation was backed up by a significant increase in safety in terms of MTTC scores.

We observed the safest behaviour in terms of MTTC scores and perceived workload in conditions where the machine's sensory capabilities were unaffected by the environment and a precise direction of safety-relevant traffic participants was provided to the user. We interpret from this that performance is linked to the availability of information. It confirms that the specific signal was appropriately used by participants to acquire a more accurate understanding of the direction of surrounding objects.

Attention distribution. Analysis of the eye tracking data revealed that visual attention was affected by the uncertainty signaling in scenarios where the human visibility was limited. The results suggested that the onset of a signal shifted visual attention towards the direction of the object that caused the uncertainty signaling.

Acceptance. Usefulness and satisfaction of the assistance were rated positively in conditions where the human and machine were uncertain and the machine provided uncertainty-encoded directional cues. We did not find significant differences when comparing these ratings with the ratings of a system that was able to sense and communicate precise directions of surrounding traffic. This is a promising observation that suggests that users are still satisfied with the directional cues and recognize the usefulness of the signals, despite the lower quality in terms of information specificness.

We found indications that the acceptance of uncertainty communication may be context-dependent; when the machine was uncertain while the driver was confident, the usefulness-satisfaction ratings showed neutral ratings. This could mean that there is less need for the machine to continuously share observations when the human agent is confident. For successful human-machine cooperation or teaming, a human mental representation of system uncertainty may not be enough; when the machine also has a representation of human confidence in different environments, it allows the machine to decide under what conditions to provide suggestions to the user. However, such a selective and presumably personalized communication could induce confusion when violating a user's assumptions on what the machine is communicating. In this example it might not even be possible for a user to unambiguously distinguish between cases in which the machine is not providing stimuli because it has not detected a potential collision event and cases in which it has selectively disabled communication because it could confirm that the user has a sufficient scene understanding. Selectively deactivating systems that implicitly encode the absence of issues through an absence of stimuli could therefore be problematic.

Types of uncertainty communication. Subjective reports seem to agree that the machine communicated uncertainty (Table 4, q5). An important difference between earlier studies that have demonstrated successful communication of uncertainty (e.g. [2, 18, 23]) is that we currently relied on an *implicit* representation of uncertainty; the uncertainty component was encoded within the spatiotemporal signaling functionality of the belt. We argue that the distinction between *implicit* and *explicit* uncertainty communication may be useful for the future design of reliability displays. Implicit uncertainty communication is characterized by an increase ambiguity or vagueness, or a decrease in specificness of presented information to convey increased uncertainty. One example of implicit uncertainty communication that we encountered in literature is by Finger and Bisantz [5], who added distortions to an image to make it increasingly difficult to specify the underlying image, thus invoking uncertainty about the contents of the image. The currently presented study falls in the same category; instead of explicitly stating that 'I am uncertain', the machine agent communicates uncertainty by being less specific in its suggestions about the location of objects.

Limitations and challenges. A limitation of the current study is that the sample may not be representative to the population (most technically schooled, 13/14 male).

We have currently only collected evidence that the system is effective in extremely challenging traffic situations, indicated by two participants as feeling unrealistic. An advantage of the rapid succession of safety-critical situations is that it allowed us to collect data with a limited amount of participants and that it ensured that the participants were exposed to the functionality of the device, which only provides stimuli when operating outside of safety bandwidth. This means that, in safe conditions, the system does not produce any stimuli. The fact that the system proved its usefulness in challenging situations can be seen as a strength, however we do not know if the effects remain when the system is not activated often. Future work could address this issue by implementing easier scenarios where a participant encounters only a few, or only one safety-critical event(s).

This study has shown that car driving is a promising application of the presented interface. Other domains that may benefit from displays that can share spatial uncertainty information may be (motor)cycling or more distant domains such as the control of ships, drones or unmanned vehicles, immersive gaming, assisting the visually impaired, (virtual) projectile impact/origin estimation or other dynamic situations where sensory capabilities may vary, or where acting forces are not fully determined.

A VIDEO

As a supplement to this paper, a video is included that shows two trials from the MU-HU-uc condition. A vehicle appears from the rear and overtakes over the left lane. The uncertainty signaling disappears as the machine concludes that the approaching vehicle is not on a collision path. In the second trial, a vehicle appears from the front and is overtaken by the ego-vehicle. The signal converges from the uncertainty signal (Figure 3) into the precise signal (Figure 2) as soon as the machine obtains a reliable estimate of the precise location of the object.

The visualization shown in the video is only a representation of what participants perceived through the belt. This visualization was not visible to the driver, only to the experimenter who could use it to confirm functionality during the experiment.

REFERENCES

[1] Audi. 2018. A8 Owner's Manual.

[2] Johannes Beller, Matthias Heesen, and Mark Vollrath. 2013. Improving the Driver-Automation Interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 55, 6 (dec 2013), 1130–1141. https://doi.org/10.1177/0018720813482327

[3] Alvaro Cassinelli, Carson Reynolds, and Masatoshi Ishikawa. 2006. Augmenting spatial awareness with Haptic Radar. In *2006 10th IEEE International Symposium on Wearable Computers*. IEEE, 61–64. https://doi.org/10.1109/ISWC.2006.286344

[4] Paulo G. de Barros and Robert W. Lindeman. 2013. Performance effects of multisensory displays in virtual teleoperation environments. In *Proceedings of the 1st symposium on Spatial user interaction - SUI '13*. ACM Press, New York, New York, USA, 41. https://doi.org/10.1145/2491367.2491371

[5] Richard Finger and Ann M. Bisantz. 2002. Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomics Science* 3, 1 (jan 2002), 1–25. https://doi.org/10.1080/14639220110110324

[6] Frank Ole Flemisch, Klaus Bengler, Heiner Bubb, Hermann Winner, and Ralph Bruder. 2014. Towards cooperative guidance and control of highly automated vehicles: H-Mode and Conduct-by-Wire. *Ergonomics* 57, 3 (mar 2014), 343–360. https://doi.org/10.1080/00140139.2013.869355

[7] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (oct 2006), 904–908. https://doi.org/10.1177/154193120605000909

[8] Sinan Hasirlioglu, Alexander Kamann, Igor Doric, and Thomas Brandmeier. 2016. Test methodology for rain influence on automotive surround sensors. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2242–2247. https://doi.org/10.1109/ITSC.2016.7795918

[9] Tove Helldin, Göran Falkman, Maria Riveiro, and Staffan Davidsson. 2013. Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13*. ACM, ACM Press, New York, New York, USA, 210–217. https://doi.org/10.1145/2516540.2516554

[10] Jean-Michel Hoc. 2001. Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human-Computer Studies* 54, 4 (apr 2001), 509–540. https://doi.org/10.1006/ijhc.2000.0454

[11] S. Hoffmann, H. P. Krüger, and S. Buld. 2003. Vermeidung von Simulator Sickness anhand eines Trainings zur Gewöhnung an die Fahrsimulation. *VDI Berichte* 1745 (2003), 385–404.

[12] J.H. Hogema, S.C. De Vries, J.B.F. Van Erp, and R.J. Kiefer. 2009. A Tactile Seat for Direction Coding in Car Driving: Field Evaluation. *IEEE Transactions on Haptics* 2, 4 (oct 2009), 181–188. https://doi.org/10.1109/TOH.2009.35

[13] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct (UbiComp '14 Adjunct)*. ACM Press, New York, New York, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

[14] Gary Klein, D.D. Woods, J.M. Bradshaw, R.R. Hoffman, and P.J. Feltovich. 2004. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems* 19, 06 (nov 2004), 91–95. https://doi.org/10.1109/MIS.2004.74

[15] Matti Krüger, Heiko Wersing, and Christiane B. Wiebel-Herboth. 2018. Approach for Enhancing the Perception and Prediction of Traffic Dynamics with a Tactile Interface. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18*. ACM, ACM Press, New York, New York, USA, 164–169. https://doi.org/10.1145/3239092.3265961

[16] Matti Krüger, Christiane B. Wiebel, and Heiko Wersing. 2017. From Tools Towards Cooperative Assistants. In *Proceedings of the 5th International Conference on Human Agent Interaction - HAI '17*. ACM, ACM Press, New York, New York, USA, 287–294. https://doi.org/10.1145/3125739.3125753

[17] Alexander Kunze, Stephen J. Summerskill, Russell Marshall, and Ashleigh J. Filtness. 2018. Preliminary Evaluation of Variables for Communicating Uncertainties Using a Haptic Seat. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18*. ACM, ACM Press, New York, New York, USA, 154–158. https://doi.org/10.1145/3239092.3265959

[18] Alexander Kunze, Stephen J. Summerskill, Russell Marshall, and Ashleigh J. Filtness. 2019. Automation transparency: implications of uncertainty communication for human-automation interaction and interfaces. *Ergonomics* just-accepted (feb 2019), 1–16. https://doi.org/10.1080/00140139.2018.1547842

[19] John D Lee and Katrina A See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (jan 2004), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

[20] Brittany E. Noah, Thomas M. Gable, Shao-Yu Chen, Shruti Singh, and Bruce N. Walker. 2017. Development and Preliminary Evaluation of Reliability Displays for Automated Lane Keeping. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '17*. ACM, ACM Press, New York, New York, USA, 202–208. https://doi.org/10.1145/3122986.3123007

[21] Jinke D. Van Der Laan, Adriaan Heino, and Dick De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5, 1 (feb 1997), 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

[22] Jan B. F. Van Erp. 2002. Guidelines for the use of vibro-tactile displays in human computer interaction. In *Proceedings of eurohaptics*, Vol. 2002. 18–22.

[23] Philipp Wintersberger, Dmitrijs Dmitrenko, Clemens Schartmüller, Anna-Katharina Frison, Emanuela Maggioni, Marianna Obrist, and Andreas Riener. 2019. S(C)ENTINEL. In *Proceedings of the 24th International Conference on Intelligent User Interfaces - IUI '19*. ACM Press, New York, New York, USA, 538–546. https://doi.org/10.1145/3301275.3302332

Appendix A

Design of a vibrotactile uncertainty communication device

This chapter provides further information about the design process of the uncertainty communication device. This chapter may serve as a reference for future designs of uncertainty communication systems or as a reference to improve reproducibility of the experiment described in Chapter 1.

The goal of the design was to create an intuitive vibrotactile representation of uncertainty about spatial observations. A method for communicating precise spatiotemporal information about surrounding objects was designed by Krüger et al. [4]. The source code and hardware (Figure A.1) from that project were available for the current project.

I will discuss three design variants for this system that may each represent spatial uncertainty. Before introducing the design variants, I summarize findings from the accompanied literature review (see Appendix G) that were useful in the design of the uncertainty communication signals:

- Uncertainty can be split up into three classes. The first class is measurement uncertainty, the second class is prediction uncertainty, and the third class is representation uncertainty. In the current experiment, we modulated the reliabiliy of the vehicle's sensors. Therefore, the most relevant class to the current design is measurement uncertainty. For the current design, we assume that measurements of the lateral position of other road users is a set of measurements with varying spread that can be visualized with a distribution function, as in Figure A.3a.
- Most approaches have focused on the visual modality. A sparsely investigated modality for conveying uncertainty information is the somatosensory receptory field.

Figure A.1: FeelSpace NaviBelt. Image provided by Feelspace.

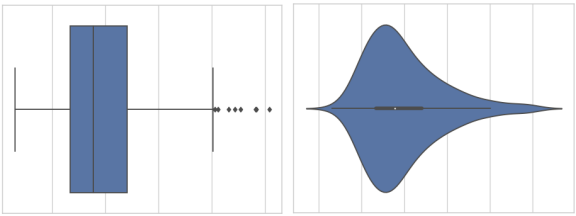


Figure A.2: Boxplot and violin plot.

- Popular visual encodings for representing measurement uncertainty seem to be the addition of vagueness, ambiguity, or spread to the signal. All these concepts are similar in the sense that they decrease specificness of information. This decrease in specificness is usually inevitable since lower-quality measurements generally have more spread. However, some authors have modulated vagueness as a parameter that could be varied regardless of measurement quality to communicate system uncertainty.

Examples where spread of a signal are summarized in a visual representation are boxplots or violin plots (Figure A.2). The boxplot and violin plot give an intuitive visual representation of a set of measurements and properties of the underlying distribution of the data set. For the design of a representation of uncertainty using vibrotactile communication, I created three representations that were inspired by these visual representations.

To explain the design of the three vibrotactile signal variants, I introduce an example situation in Figure A.3a. The ego vehicle obtains measurements about the lateral position an object that is approaching from the front. The distribution of these measurements is visualized on the top. It may be regarded as a probability density function of the actual position of the object.

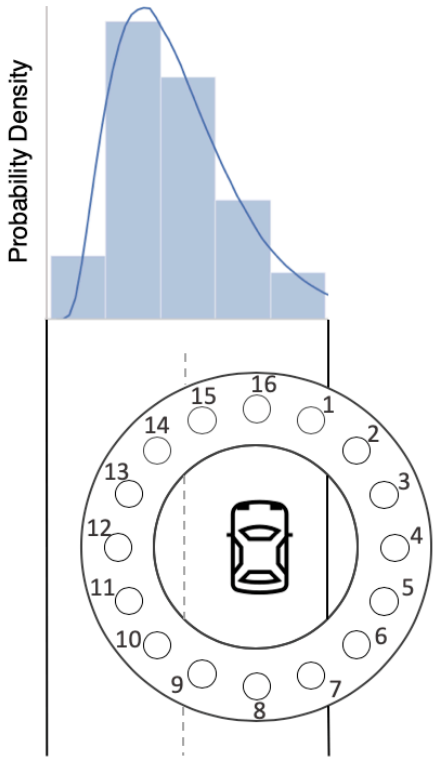
A.1 Signal variant 1

An early design iteration included the information about the underlying distribution of the data into the vibrotactile signal. Figure A.3b shows the activation intensities of vibromotors in the waist belt for this case. It can be seen that the vibromotors that lay in the direction of the front two lanes (motors 13 - 1) are activated at intensities that directly represent a dicretization of the underlying distribution function from Figure A.3a.

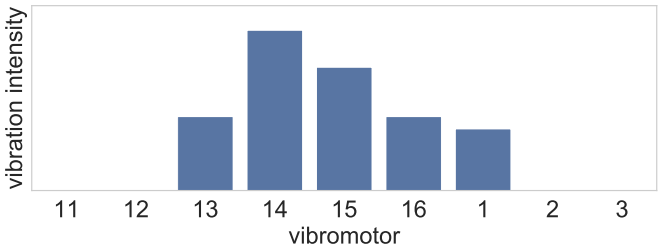
Experimenter tests with this signal representation revealed a couple of shortcomings. It was nearly impossible to reconstruct the shape of the distribution from the feeling that the waist belt induced. We noticed that the same shape felt different when presented on different parts of the body. This is likely caused by variance in proprioceptive capabilities of the somatosensory receptive field around the human waist.

Another cause of this problem may be the low spatial resolution of the vibrators on the waist. We have a resolution of 16 vibromotors per 360 degrees of vision. When communicating about objects on a two-lane road, as in the example situation, we could effectively use five vibromotors to encode the shape of the distribution function. A belt with more vibromotors may be more suited to convey the shape of the distribution in a vibrotactile signal.

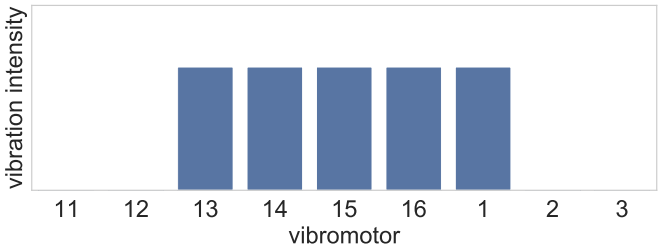
Another problem was that the generation of a shape required the stimulation intensity of the vibro-



(a) Schematic representation of measurements when an object approaches the ego vehicle from the front.



(b) Signal variant 1.



(c) Signal variant 2.

Figure A.3

motors to be modulated individually. In the design by Krüger et al. [4], the stimulation intensity was a function of TTC. It was too difficult to combine representations of a shape of the distribution function and an indication of the TTC variable.

A.2 Signal variant 2

As a second design alternative, the shape of the underlying distribution was removed. Instead, a range of vibromotors was activated with equal stimulus intensity (Figure A.3c). This resulted in a stimulus of variable width (i.e. amount of vibromotors activated). A wider signal corresponded to measurements with less certainty (more spread) and a narrow signal corresponded to a precise estimate of the position of an object.

The intensity of the signal was a function of TTC with the approaching object, with the same TTC-intensity relation as described in Chapter 1 (section *Signal For a Confident Machine*). Since there was no need to encode the shape of an underlying distribution with the intensity of the signal, this approach seemed more promising.

However, there were still some issues with this signal. As more vibromotors were activated for measurements with more spread, this resulted in more perceived intensity of the signal. This was an unwanted effect, since we intended to keep perceived intensity unchanged when the TTC did not change.

A.3 Signal variant 3

A third signal variant was eventually used in the experiments. Signal variant 3 is described in Chapter 1 (*Signal for an Uncertain Machine*). This signal variant seemed promising as we could distinguish small changes in widths of the signal, whilst keeping the ability to understand information through the vibration intensity parameter. This meant that it was a promising method to use in combination with a TTC encoding on the vibration intensity parameter.

Illusion of motion

The uncertainty signaling device made use of the apparent motion illusion. The haptic variant of this phenomenon was first described in 1917 by Burtt [1], who found that an illusion of movement between two actuators can be induced by sequentially presenting two tactile stimuli in close proximity. Guidelines for creating reliable continuous vibrotactile movement sensations and patterns can be found in [6] and [3].

Design parameters of the uncertainty signal

The uncertainty signal (signal variant 3) as used in the experiment conveyed a measure of uncertainty which was based on measurement spread, as well as a measure of spatiotemporal proximity (TTC). This subsection describes details of the uncertainty signal.

In the experiment, the switch between a certain and an uncertain state was simulated as a discrete event, namely when an object was more than 33 m away from the ego vehicle (when driving

in sub-optimal machine sensory conditions like rain or a tunnel). However, more realistically, measurement uncertainty is based on a continuous measure of signal spread such as variance or the width of a confidence interval.

The uncertainty measure was communicated by displaying a dynamic pattern that traveled back and forth between two boundaries. The steps to creating the uncertainty signal are listed below, see Figure A.4 for accompanying illustrations.

1. A measurement of the angular position of an object (between 0 and 360 degrees) is represented here as a probability density function of the position of an object. Borders b1 and b2 are constructed by finding an area C that equals a pre-defined confidence level. Naturally, for measurement distributions with less spread, the distance between b1 and b2 will decrease.

2. Actuators positioned between the angle coordinates b1 and b2 will present the oscillating signal. The range of actuators within b1 and b2 make up the *oscillation path*.

3. The oscillating pattern is initialized by selecting an inital actuator on the oscillation path. In the experiments, it was selected at random but it may also be based on the shape of the distribution function (e.g. the position with highest likelihood). The initial actuator starts vibrating for a *vibration time* T_{vib}, after which it is followed up by its neighbour. The initial direction of the oscillating signal was also selected at random.

4. The *oscillation time*, the time it took for the signal to travel up and down the oscillation path, was set to 1.0 s. This means that the vibration time T_{vib} was dependent on the number of actuators within the oscillation path n: $T_{vib} = 1.0/n$.

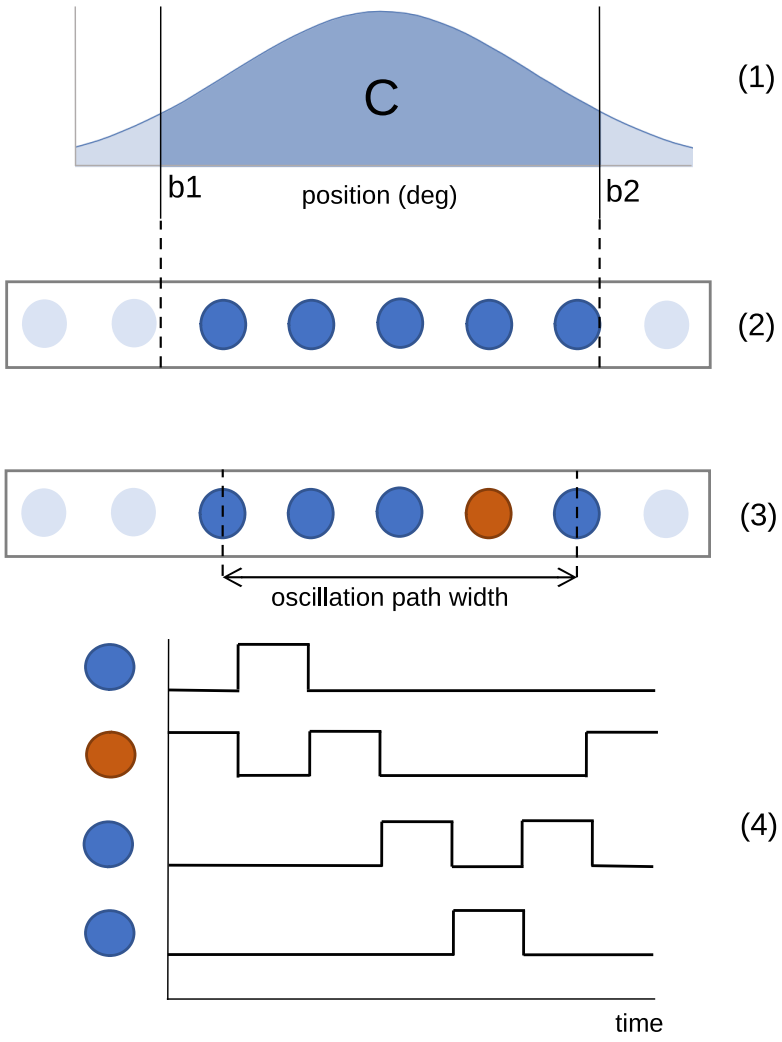


Figure A.4: Steps to generation of a signal.

Appendix B

Design of the Experimental Conditions

This section gives a more detailed description of the implementation of the scenarios in the Silab environment. In Table B.1 the Silab settings that were used to create the different visibility conditions (e.g. time of day, air humidity) are listed. A detailed description of the traffic profile for foggy scenario's is discussed first, followed up by a scenario that was designed for testing

B.1 Traffic profile for foggy scenarios

As was explained in the paper, approaching vehicles were launched by the experimenter upon a button press. How this was implemented is explained in this subsection.

Figure B.1 shows a schematic overview of the initial traffic setup in the foggy scenarios. Twelve cars were placed at a locked distance of 110 meters m in front and behind the ego vehicle, distributed equally over the two lanes and in front and behind the Ego vehicle, in groups of three. In addition, two more vehicles were placed on the right lane, locked at a fixed distance of 400 meters m in front of the ego vehicle. All vehicles were located well outside the visible range of the driver and their velocity was set to always match the velocity of the Ego vehicle. This meant that these groups of vehicles kept driving at the same, locked distances at which they were set up. The individual vehicles within a group were driving on the exact same coordinates; the Silab

Table B.1: Silab parameters

		Scenario Type		
Parameter	**Silab Class**	**Familiarization**	**Rain**	**Fog**
World rendering mode	SGEWorld.Rendermode	8	8	2
Time of day	Environment.InTimeOfDay	14.00	14.00	3.00
Air humidity	Environment.InHumidity	0.50	0.50	1.05
Rain intensity	Environment.InPrecipitation	0.00	0.25	0.00

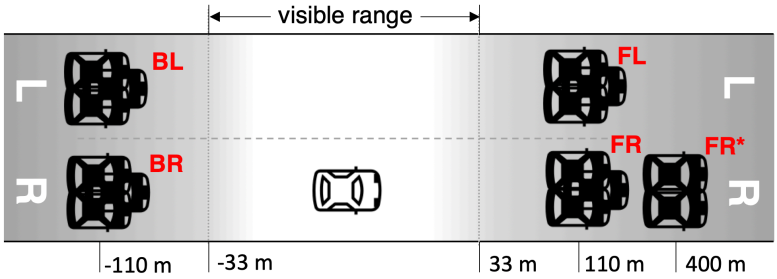


Figure B.1: Traffic in scenarios with fog (human uncertain) scenarios.

Table B.2: Order of launched vehicles. Vehicles annotated with an * approached at 0 km/h instead of 80 km/h. Due to an experimental design error they were excluded from the analysis (see Section B.2).

1	2	3	4	5	6	7	8	9	10	11	12	13	14
FL	FR	FL	BL	BR	FR*	FL	BL	FR*	BR	BL	BR	FR	FR

simulator software allowed vehicles to drive trough each other. The vehicles simply behaved as ghosts passing through each other.

Vehicles would keep driving in their groups unless a vehicle was unlocked by a key press commanded by the experimenter. After such a command, a vehicle instantly decreased its speed to 80 km/h (when approaching from the front), or increase its speed to 160 km/h (when approaching from the rear). The two additional vehicles that were placed at a 400 m distance would not slow down to 80 km/h, but instead come to a full stop. As a consequence of its new velocity, the surrounding vehicle would pass (or collide with) the ego vehicle shortly after the experimenter gave the unlocking command (assuming the ego vehicle kept driving at a velocity of around 120 km/h). After a vehicle passed and disappeared into the fog, and the experimenter confirmed that the participant had reached a velocity of approximately 120 km/h again, the next vehicle was launched. This was repeated until all 14 vehicles were launched.

Cars that approached from the front right lane could be avoided by the driver by overtaking over the left lane. Cars originating from the front left could be overtaken over the right lane. From the rear end, the vehicles overtook the ego vehicle via the left lane. The rear end cars that started out on the right lane would first accelerate to 160 km/h, and then blink and switch to the left lane to overtake at a distance of 30 meters.

The direction from which cars approached was a randomized sequence, re-used in each experiment for comparability. The order of the encounters is listed in the paper and in Table B.2.

The purpose of launching the vehicles individually from behind a fog field was to create a series of controlled trials in a relatively easy way. Since the driver had no way of knowing what was going on behind the fog curtain, there was no need to implement realistic traffic behavior behind this curtain. No complex driver behavior models or choreographies had to be implemented for the surrounding vehicles. From the driver's perspective, it only seemed like cars were appearing out of the fog and disappearing into it again.

B.2 Design error for foggy scenarios

Two vehicles that approached from the right lane were completely standing still (FR*, Table B.2). These were placed further away from the ego vehicle (400m) than the three vehicles that launched with 80 km/h. These two extra vehicles were included in the experiment since we wanted to have different levels of difficulty in evading the oncoming traffic.

We found out that when the standing-still cars were triggered, there were were often still vehicles driving between them and the ego vehicle. In the current implementation of our algorithm, only the nearest vehicle was considered for communication. So, when an FR* vehicle was launched, it would be signalled to the user only after it had passed the FR group. This group was passed at a distance of 110 m, whilst the relative velocity between the ego and the FR* vehicle was 33 m/s, meaning that the TTC between the ego vehicle and the FR* vehicle at the moment it was first signalled, was $110/33.33 = 3.0$ s. This is far below the signalling threshold that was intended (9.0 s).

For this reason, FR* vehicles were not comparable to each other since they were not signalled to the user at comparable time instances. Unfortunately, this problem was identified after the experiments took place. We excluded them from the analysis.

Appendix C

Data Analysis

This chapter provides further details of the data analysis process. I will first describe how participant data was structured and how it was split up in smaller time fragments. Then I will follow the structure of the Results section of the paper (Chapter 1) to explain how the results were obtained from the data. The jupyter notebooks are stored in the digital supplement (see Appendix D).

C.1 Software

Python 3.7 was used for the data analysis. For handling the data structures I relied on the pandas package (v. 0.23.4) and for statistical analyses on SciPy's stats package (v. 0.14.0). Figures were produced using Matplotlib (v. 3.0.3) and Seaborn (v. 0.9.0).

C.2 Slicing up the recordings

As was explained in Chapter 1, individual recordings were sliced up into smaller time fragments. The start and ending points of a time fragment were dependent on the time to passing (TTP) of surrounding objects. In this chapter, I explain why the TTP variable was used to determine the edges of a time fragment and how the extraction of the time fragments worked.

C.2.1 Why TTP?

We were interested in the changes in user behaviour that were caused by the onset of a communication signal. In conditions with full machine overview (MC-HU), the machine signaled the precise signal for any vehicle that was (1) on a collision path with the ego-vehicle and (2) the TTC of the collision was under 9.0 s. In the conditions where uncertainty communciation functionality was enabled (MU-HU-uc and MU-HC-uc), the machine was uncertain about the lateral position of objects further then 33 m away from it. In other words, it had no information to conclude on what lane an object was driving. In these cases, the machine presented the uncertainty signal for objects driving on either of the two lanes, as long as they had a potential TTC of 9.0 s. The word potential is used here, since there did not need to be a direct collision trajectory between the object and the ego vehicle for the uncertainty signal to be triggered. This is illustrated in Figure X. This

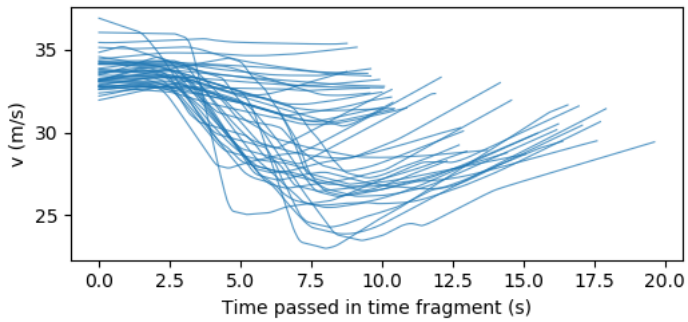


Figure C.1: Velocity development of time fragments in MU-HU-uc condition.

potential TTC is the time it would take until both vehicles are on the same height (longitundinal position) as each other, which is the same as the time to passing (TTP).

It is important to note that the starting point of a time frame is not calculated by taking the time stamps at which vehicles passed the ego vehicle and subtracting 9 seconds from it. Like TTC, TTP is a dynamic variable that can be calculated at any point in time. This means that time frames are not of equal length. Participants that decided to slow down after the onset of a signal will have longer time frames then participants that maintained their current velocity and switched lanes. This is apparent in Figure C.1, where the velocity during each time fragment is plotted.

C.2.2 Extracting the time stamps from a recording

Here I demonstrate how the start and ending moment of a time fragment were determined from a full recording.

Figure C.2 shows the development of TTP during a recording in MU-HU-uc (foggy tunnel) conditions. For this explanation, I only show the TTP for vehicles that approached from the front. From the figure, we can identify eight instances where the TTP equaled zero. Each of these time instances represent a moment a vehicle passed the driver. These were directly used as the ending points of the isolated time fragments. The starting points were determined by finding the intersection points with the TTP = 9 s line and selecting the intersection points that preceded an end point. These intersection points are visualized with round bullets.

With this method we can extract the beginning and ending time stamp of the time fragments. By extracting the data between these time stamps, we generated the raw data visualizations as in the paper (Chapter 1).

C.3 Analyses

In this section I provide more detailed descriptions of the statistical analyses that were conducted in the paper. I will justify my choice of statistical tests for the different obtained measures. I will follow the structure of the paper, and for ease of explanation the figures from the paper are copied to this section.

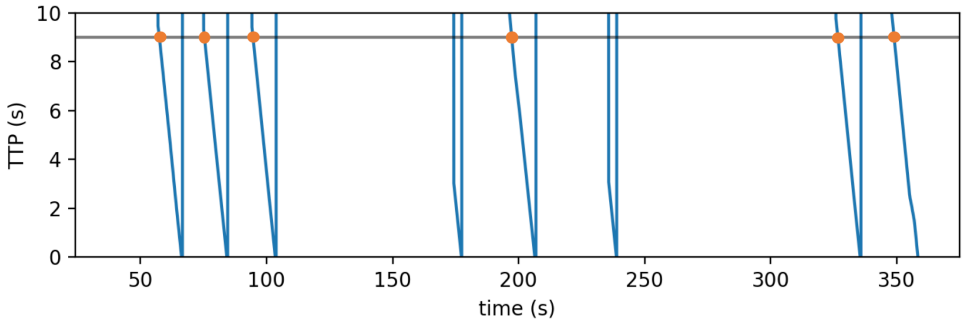


Figure C.2: Development of TTP for one full recording in a MU-HU-uc condition. Note that the two approaches at t = 177 s and t = 239 s are preceded by an intersection point with the TTP = 9 s line, however there is no orange marker to indicate this. These peaks correspond with the two vehicles that stood completely still, where the experimental design error as discussed in section B.2 was made. These approaches were excluded from the analysis.

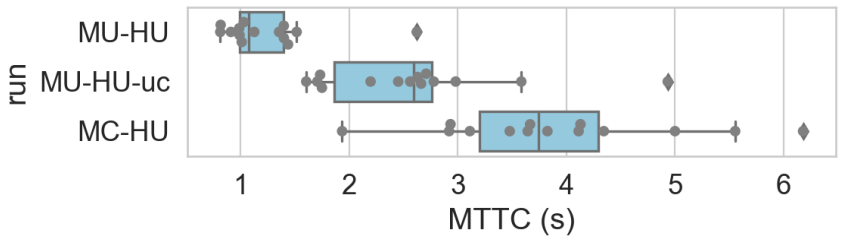


Figure C.3: MTTC scores for human uncertain conditions (n = 14).

C.3.1 MTTC

The MTTC scores for the human-uncertain conditions are summarized in Figure C.3. It seems apparent that the MTTC scores for the three conditions are different. We wanted to back this figure up with inferential statistics. In particular, we wanted to compare the system with uncertainty communication (MU-HU-uc) with two other conditions. First, a comparison was made with a system that was not capable of expressing uncertainty but instead communicated nothing until it had a reliable estimate of the position of an object (MU-HU). Secondly, a comparison was made between a system that had a reliable sensory input, meaning that it was able to obtain accurate measurements of the position of objects that were further away (>33 m) and that communicated the precise signal for these vehicles.

We expected that in the first comparison, the uncertainty communication condition would have higher MTTC scores, indicating earlier takeovers or earlier braking.

In the second comparison, the expectation was that the precise communication condition (MC-HU) would have safer (higher MTTC) scores than the condition with uncertainty communication. If the precise signaling function was interpreted by the participants as the design was intended, the participant can infer on what lane an object is driving before he/she sees the object and change to the unoccupied lane earlier than when receiving an uncertain signal, that only signals *that* there is something but not on what lane that object is.

Since we have a repeated measures experimental design, paired statistical tests were used. The distribution of the data was inspected visually with a histogram and a QQ-plot [2]. In a QQ-plot, the quantiles of the obtained data are plotted against the theoretical quantiles of a normal distribution. Samples drawn from the same distributions would approximately lie on the line y=x.

Figure C.4 shows these plots for both comparisons. Note that we did not plot the distribution of the MTTC scores for each group, but instead the difference between the paired MTTC scores per participant in both comparisons. To justify the use of the paired t-test, the sample distribution of these differences have to be normally distributed. In Figure C.4c, a QQ-plot of 14 samples drawn from a normal distribution is shown for comparison. After a visual inspection of the histogram and QQ plot for our data, we assumed that the sampling distributions were approximately normally distributed. The paired t-test was used. Results can be found in Chapter 1.

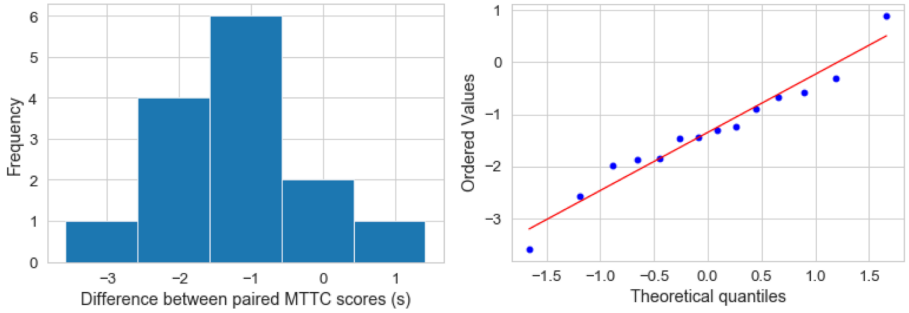
Software used

Figure C.3 was created using the Python 3.7 function `seaborn.boxplot` and overlaying it with a categorical scatter plot created with `seaborn.swarmplot`. Histograms were created with the `matplotlib.pyplot.hist` and QQ-plots using `scipy.stats.probplots`. The paired samples t-test (Chapter 1) was performed with `scipy.stats.ttest_rel`.

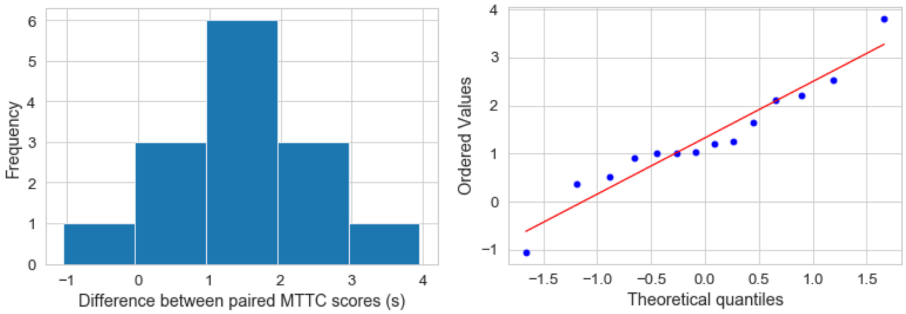
C.3.2 Eye Tracking

For eye tracking we used hardware (Figure C.5) and software from the open-source eye tracking platform by Pupil Labs [5]. The software allowed us to specify pre-defined areas of interest that were mapped to the real world using 2d markers projected on the screen (Figure C.6).

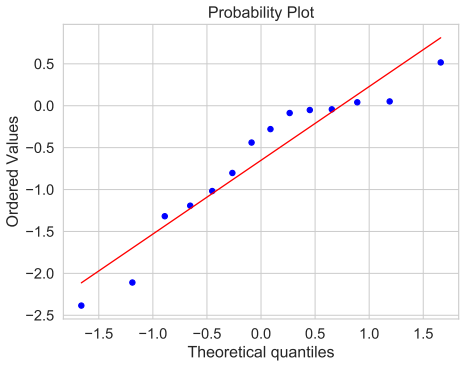
Surfaces were defined around the front window, instrument cluster and the three side mirrors. For the mirrors, we defined areas that were bound tight around the edges and one area that contained



(a) Paired score difference between MU-HU and MU-HU-uc (n=14)



(b) Paired score difference between MC-HU and MU-HU-uc (n=14)



(c) QQ plot for 14 samples drawn from a normal distribution.

Figure C.4: Histograms and normal QQ-plot of the differences between paired MTTC scores for the two comparisons.

Figure C.5: Eye tracker by Pupil Labs. Source: Pupil Labs Docs [5] (GNU LGPL licence).

Table C.1: Gaze distribution example. 'optimist' surfaces refer to the wider drawn surfaces.

	Surface Name	Count
1	topmirr	1541
2	topmirr_optimist	2885
3	leftmirr	78
4	leftmirr_optimist	166
5	frontwindow	19097
6	instrument_cluster	1598
7	not_on_any_surface	4077

the mirror plus an extra 15 cm of margin around it (see Figure C.6). These were drawn wider around the mirrors since we noticed that the estimated gaze position of participants susceptible to drift. One cause for this could have been that the eye tracking glasses slightly moved on the participants head during experiments.

Between specified start- and ending points, the eye tracking software calculated a total number of identified gaze points on each surface. An example of a gaze distribution can be seen in Table C.1. Such a distribution was created for each isolated time fragment.

The total counts are dependent on the length of a time slice and on the quality of a recording: longer time slices will have higher counts in general, and poorer-quality recordings contain lower counts. To counter these effects, I chose to evaluate the ratios within a gaze distribution. More specifically, we decided to only look at the ratio of counted gaze points towards the front window and counted gaze points towards any of the mirrors plus the front window. This resulted in a number between 1 (only gazed through front window) and 0 (only gazed through mirrors). For the mirrors, the optimistically drawn surface areas were used.

Results of the analysis can be found in the paper (Chapter 1). For the statistical analysis of the gaze ratio the python function `scipy.stats.wilcoxon` was used to perform the Wilcoxon signed rank test.

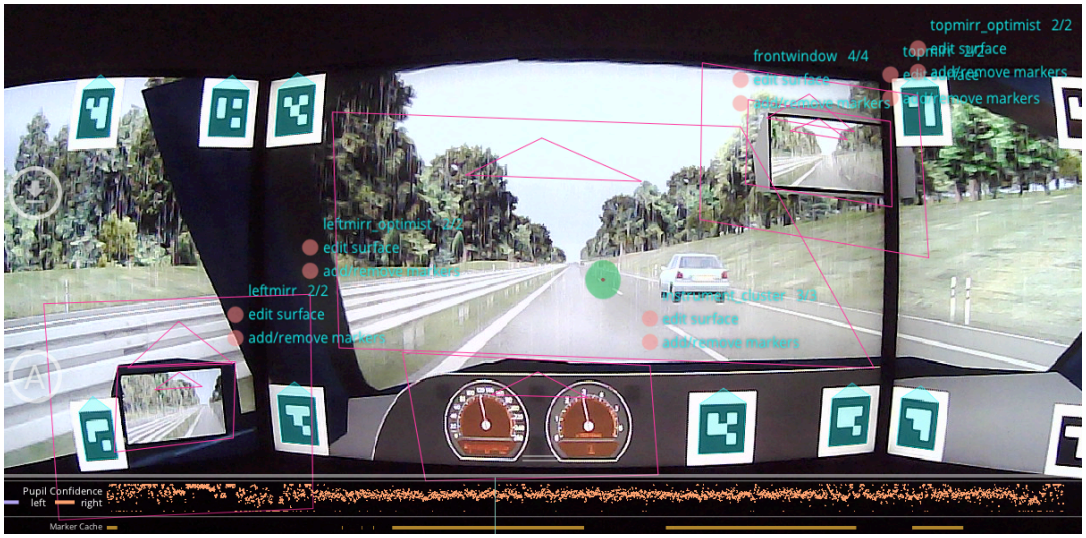


Figure C.6: Screen shot from the eye tracking software. Pink outlines show the augmented surfaces mapped to the black and white markers. The green circle in the middle indicates the currently estimated gaze point.

Bibliography of Appendices A-C

[1] Harold E. Burtt. “Tactual illusions of movement.” In: *Journal of Experimental Psychology* 2.5 (1917), pp. 371–385. ISSN: 0022-1015. DOI: 10.1037/h0074614. URL: http://content.apa.org/journals/xge/2/5/371.

[2] John M. Chambers et al. *Graphical Methods for Data Analysis*. Chapman and Hall/CRC, Jan. 2018. ISBN: 9781351072304. DOI: 10.1201/9781351072304. URL: https://www.taylorfrancis.com/books/9781351080750.

[3] A. Israr and I. Poupyrev. “Control space of apparent haptic motion”. In: *2011 IEEE World Haptics Conference*. IEEE, June 2011, pp. 457–462. ISBN: 978-1-4577-0299-0. DOI: 10.1109/WHC.2011.5945529. URL: http://ieeexplore.ieee.org/document/5945529/.

[4] Matti Krüger, Heiko Wersing, and Christiane B Wiebel-Herboth. “Approach for Enhancing the Perception and Prediction of Traffic Dynamics with a Tactile Interface”. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI ’18*. ACM. New York, New York, USA: ACM Press, 2018, pp. 164–169. ISBN: 9781450359474. DOI: 10.1145/3239092.3265961. URL: http://dl.acm.org/citation.cfm?doid=3239092.3265961.

[5] *Pupil Labs Eye Tracking Software*. URL: docs.pupil-labs.com (visited on 05/13/2019).

[6] Jan B. F. Van Erp. “Guidelines for the use of vibro-tactile displays in human computer interaction”. In: *Proceedings of eurohaptics*. Vol. 2002. 2002, pp. 18–22.

Appendix D

Digital Supplement

The digital supplement of this thesis is available online at
`https://github.com/tomdries/feelinguncertain`

D.1 Analysis

The analysis was performed with python, implemented in Jupyter notebooks. The notebooks can be found in the `Analysis` folder in the digital appendix. The notebooks were saved as .html files, so they can be opened directly in any browser. The notebooks are also stored in their original form, in the underlying `code` folder. These can be executed using Jupyter.

D.2 Example of Raw Data

Raw data for a test run by one of the experimenters can be found in the `Rawdata_example` folder. The recordings contain Silab recording files and Python recording files (.csv). Silab datafiles contain measurements directly obtained from the car simulator software. This data was transferred to a second computer that ran a Python script transforming this data into variables required for the uncertainty signal (Appendix A). The Python script controlled the waist belt. Tables D.1 and D.2 provide a description of the recorded Silab and Python variables, respectively. Due to privacy regulations, we did not publish the recorded raw data from the experiment.

D.3 Videos

The `Videos` folder contains three files. All videos were recorded during a test run where one of the experimenters was in the driver seat. Video A shows example situations that occured in the MUHUuc condition. A visualization of the current actuation via the belt is displayed. This visualization was also available live, but only to the experimenter. Video B shows situations in the same condition for situations where the approaching vehicles were completely standing still. These vehicles were excluded from analysis due to the experimental design error explained in section B.2.

Table D.1: Silab recording file variables. For items denoted with a *, a naming convention is needed: the abbreviation SAN describes the lane direction Same lane, Ahead of ego, Next vehicle, SBN denotes Same lane, Behind ego, Next vehicle. L and R denote an offset to the left and right lane. So, for example, StLongSANL is the longitudinal distance to the nearest vehicle driving ahead of Ego, left of Ego's current lane.

Name	Description
MeasurementTime	Time (ms)
Measurement time error	Time lag due to system overload (ms)
AccelPedal, BrakePedal	Acceleration and Brake pedal depression [0,1]
SteeringWheel	Steering wheel rotation [-1, 1]
yaw	Yaw (deg)
vEgo	velocity (m/s)
vLateral	lateral velocity (m/s)
vyaw	yaw rate (deg/s)
ax, ay	Acceleration in x (lateral) and y (longitudinal) direction (m/s^2)
s	Distance traveled (m)
TLCLeft, TLCRight	Time to lane change (s)
LaneIdx	Current lane ID
vSAN, vSANR etc.*	Velocity of surrounding vehicles* (m/s)
StLongSANL etc.*	Longitudinal distance to nearest surrounding vehicles* (m)
StLatSANL etc.*	Lateral distance to nearest surrounding vehicles* (m)

Table D.2: Python recording file variables

Name	Description
angle_front, angle_rear	At what angles the actuators are currently vibrating (deg)
b1, b2	Between these borders the uncertainty signal oscillates for front objects (deg)
b3, b4	Between these borders the uncertainty signal oscillates for rear objects (deg)
muf, mur	Lane direction of most urgent vehicle from front (muf) and most urgent vehicle from rear (mur), based on lowest TTC
pBL, pBM, etc.	For all directions the p parameter is the intensity of the belt, varying between 0 and 1.
t_eye, t_silab, t_python	Timestamps for eye tracking, silab and python recordings. Used for synchronisation. t_silab corresponds to 'MeasurementTime' in silab recordings (ms)
ttcBL etc.	Computed TTC values for each lane direction. (s)
vEgo	Velocity (m/s)

Video C shows a short screen capture from the eye tracking software in a rainy scenario. This gives an impression of the implementation in the pupil labs software.

Appendix E

Questionnaires

Demographics

1. What is your age?

2. What is your gender?

☐ Male ☐ Female ☐ I prefer not to respond

3. At which age did you first obtain your driver's license?

4. What is your primary mode of transportation?

☐ Private vehicle ☐ Public transportation ☐ Motorcycle

☐ Walking/cycling ☐ Other ☐ I prefer not to respond

5. 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 then once a month ☐ Never

☐ I prefer not to respond

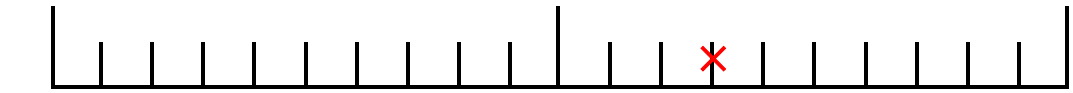
6. Did you participate in the previous driving simulator experiment conducted by Matti Krueger? The task was to drive through scenarios while receiving vibration feedback from a waist belt.

☐ Yes ☐ No ☐ I prefer not to respond / I'm not sure if I did

NASA TLX Calculation Example

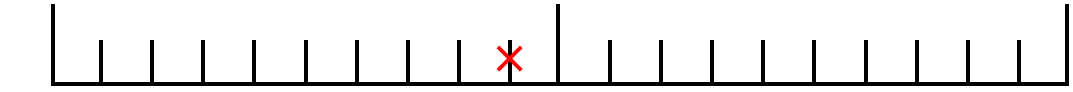
The questions below are about your experience in the run that you just performed. Put a cross on the line, not between them[1].

Mental Demand How mentally demanding was the task?



Very Low Very High

Physical Demand How physically demanding was the task?

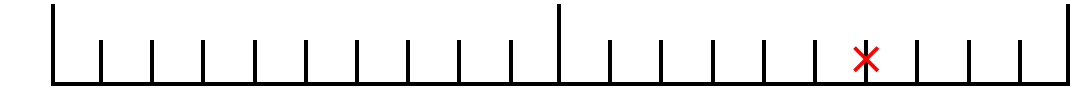


Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?



Perfect Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low Very High

All dimensions are scored on a scale 0-100%. The overall workload is then determined by the average of the six dimensions. For this example, the overall TLX is = SUM / 6 = (65+45+30+80+55+25) / 6 = 50%

Source: Kyriakidis, M., De Winter, J. C. F., Happee, R. (2014). *Human Factors of Automated Driving – Recommended Questionnaires*

[1] In the case that a participant puts the cross between the lines, the score of the line on the right-hand side should be taken.

Acceptance Scale Calculation Example

My judgments of the (...) system are ... (please tick a box on every line)

1. Useful				X		Useless
2. Pleasant				X		Unpleasant
3. Bad			X			Good
4. Nice				X		Annoying
5. Effective				X		Superfluous
6. Irritating		X				Likeable
7. Assisting				X		Worthless
8. Undesirable		X				Desirable
9. Raising Alertness					X	Sleep-inducing

The scoring is on a scale +2 .. −2.
The items 3, 6, and 8 are mirrored and should be scored −2..+2.
The *Usefulness* scale is the sum of items 1, 3, 5, 7, and 9, divided by 5 (so that it has a range from −2 to +2)
The *Satisfaction* scale is the sum of items 2, 4, 6, and 8, divided by 4.

Usefulness calculation = (−1 + 0 + (−1) + (−1) + (−2)) / 5 = −1
Satisfaction calculation = (−1 + (−1) + (−1) + (−1)) / 4 = −1

Source: Kyriakidis, M., De Winter, J. C. F., Happee, R. (2014). *Human Factors of Automated Driving – Recommended Questionnaires*

Further Evaluations

		Strongly disagree	Disagree	Neutral	Agree	Strongly Agree	No Answer
1	The other road users made me unconfident						
2	The weather conditions made me unconfident						
3	The signals from the belt made me unconfident						
4	The machine was sometimes uncertain about the exact location of a vehicle						
5	The machine told me it was uncertain about the exact location of a vehicle						
6	I relied on what I perceived with my eyes						
7	I relied on what I perceived through the belt						
8	I had trust in my own capabilities						

Appendix F

Invitation and Informed Consent Form

F.1 Invite

Sent via e-mail to Honda Research Institute EU employees and students

Dear all,

We are conducting a small driving simulator and eye-tracking study and are looking for participants.

The experiments will start tomorrow (wed 24 oct) and continue only until we have reached our target number of participants.

If you would like to participate and meet all requirements listed below, please contact me to schedule a time slot. First come, first served.

Requirements:

- Student or associate currently working at HRI
- Valid driver's license
- Normal or corrected-to-normal vision (Contact lenses: no problem. Glasses: can be fine if the frame is sufficiently thin.)
- No dark eye-makeup or other visible alterations (mascara, extended eyelashes, etc.) during the experiment
- Waist circumference approximately between 80 and 100 cm (due to equipment limitations)
- 60 minutes of your time
- According to internal regulations, taking part in the experiment is not allowed for persons that have:
- epilepsy,
- heart disease,

- pregnancy,
- severe dizziness,
- severe back or neck problems that were recently medically treated.

Please make sure you fulfil all requirements if you volunteer to participate.

The driving simulator itself is static which leads to a mismatch between visual and vestibular sensory input. For some people this mismatch can create discomfort. The experiment is designed to reduce this sensory mismatch and thus related potential discomfort to a minimum. Additionally, a simulator familiarization procedure to further prevent potential negative effects will be carried out prior to the experiment.

Please don't hesitate to contact me if you have any questions or concerns.

I am looking forward to your messages.

Best regards,

Tom Driessen

F.2 Informed Consent Form (see next page)

Honda Research Institute Europe GmbH
Carl-Legien Str. 30
D-63073 Offenbach/Main
Germany

Informed Consent Form

Project Leaders: Tom Driessen, Matti Krüger, Christiane Wiebel-Herboth
Affiliation: Honda Research Institute EU
Project Title: Effects of communicating uncertainty via the somatosensory receptive field in a simulated environment

Please read the following material carefully to ensure that you are informed about the nature of this study and about how you will participate in it, if you consent to do so. Signing this form will indicate that you have been so informed and that you give your consent.

Introduction: Purpose of the Research

In this study we want to evaluate driving behavior in a set of simulated driving scenarios.

Procedure: Tasks for the Participants

Prior to the experiment you will need to enter a "first" and a "last name" which provide the basis for an ID that is associated with your recording. The ID is used to provide the recorded data with an identity that can nevertheless not directly be associated with a particular person. The names which you enter do not have to correspond to your real name but please remember your entries to allow for possible association between data from this and possible future experiments.
Furthermore, contextual information that may potentially influence your tactile perception abilities is recorded. Usually this is information about the material and thickness of your clothing. Furthermore age, gender and driving experience information is recorded.
To get used to the driving simulator and prevent the occurrence of simulation sickness, a familiarization procedure will be carried out prior to the experiment. For this procedure you will be asked to complete a series of small tasks with the simulator. Please notify the experimenter in any cases of discomfort or uncertainty about the current task.

During the experiment you will be wearing a belt equipped with multiple vibrotactile actuators which are used to produce tactile stimuli.
In addition you will be wearing head-mounted eye-tracking device which is used to record your eye-movements and infer at which screen positions you are looking while driving through the virtual environment. The device will record your eye movements on the screen while you are doing the task and a camera on top of the device will track the scene in front of you. In order to use the eye tracker, it might be necessary to adjust the cameras such that your eyes can be tracked optimally. Before the experiment and between blocks, the eye tracker has to be calibrated. This is done by a standard calibration procedure, where you are asked to subsequently fixate the center points of nine different targets on the computer screen.

The actual experiment consists of five blocks of equal length. In each block your task will be to drive on a virtual highway while following German traffic regulations and avoiding accidents. You should furthermore try to match speed regulations whenever possible and, if available, follow visually displayed navigation instructions.
In some blocks you will receive vibrotactile stimuli during driving. The goal of these signals is to enhance your perception of the surroundings. Furthermore, the signals are capable of expressing system uncertainty when the machine is less confident about an observation or a prediction it

made. There will be a procedure to familiarize yourself with these signals.
After completion of the driving tasks you will be asked to fill out a questionnaire which contains several questions about your subjective experiences during the experiment.

Risks and side effects

For some participants the use of the driving simulator can induce simulation sickness, a form of motion sickness that can occur due to a mismatch between vestibular and visual sensory input during simulation. To prevent simulation sickness a familiarization procedure is carried out prior to the experiment. Snacks and refreshments are offered as an additional countermeasure. Please inform the experimenter in cases of any discomfort.
This study is harmless and pain free for the participant according to our present knowledge.
The eye tracking device contains infrared LEDs which shine on the pupils whenever the eye-tracker is active. These LEDs are certified according to EN 62471:2008 norm exempt group (no photobiological hazard) and are thus considered safe for eyes and skin.
The vibromotors of the belt operate below potentially painful levels and the belt as a whole has been certified as safe for consumers according to EU standards.

Personal Data Handling and Confidentiality

Data security regulations are strictly obeyed. Personal data will not be passed on to third parties. The data obtained from you will be anonymized and only processed or published in this form.

Voluntary Participation and Right to Refuse or Withdraw

Your participation in this study is entirely voluntary. It is your choice whether to participate or not. The choice that you make will have no bearing on your job or on any work-related evaluations or reports. You may change your mind later and stop participating even if you agreed earlier on without giving any reason for your decision. Also any data recorded up to that point can be deleted immediately if you wish to do so.
Note however that you can only withdraw from participation until the end of the experiment. Afterwards it won't be possible to access or delete your individual data as no personal identifier will be stored with the data.
You are furthermore asked not to reveal any details of the actual content of this study to people working at HRI-EU for a period of at most 4 weeks or until all recordings of this study have been completed in order to avoid potential biases in future recordings.

Who to Contact

If you have any questions, you can ask them now or later. If you wish to ask questions later, you may contact any of the following: Tom Driessen (t.driessen@student.tudelft.nl), Matti Krüger (matti.krueger@honda-ri.de) or Christiane Wiebel-Herboth (christiane.wiebel@honda-ri.de)

Hereby I confirm that I have read and understood the foregoing information. I had the opportunity to ask questions about it and any questions I have asked have been answered to my satisfaction. I hereby authorize the processing and analysis of the data obtained from me in this study. I also give my consent to the publication of results based on data recorded from me in an anonymized way. I consent voluntarily to be a participant in this study. I furthermore agree not to share any details of this study with colleagues for a period of at most 4 weeks or until the recordings of this study have been completed.

Name of Participant_____
Signature of Participant _____
Date _____
Day/Month/Year

Page 2 of 2

Appendix G

Literature Review

Literature Review: Human-Machine Interfaces Sharing Real-Time Measures of Machine Reliability

Tom Driessen

INTRODUCTION

In scenarios where humans and machines share responsibility for executing tasks, an appropriate understanding of the operating domain and functional limitations of the machine can be crucial for safe use. Inappropriate trust in the competences of automated systems may lead to misuse or disuse of automation [1].

In dynamic conditions, the reliability of a machine's subsystems can change [2]. For instance, systems that rely on camera vision may perform worse in the dark, or the accuracy of Lidar measurements tends to decrease in the rain [3]. Since operators cannot be expected to have an understanding of the underlying mechanisms (or the mere existence) of these subsystems, users may benefit from the availability of measures of machine reliability. A machine that shares updates about self-assessed reliability measures may help drivers to dynamically adjust their level of trust in the automation to appropriate level. Adjusting trust towards levels that are more in line with the actual reliability of other agents is known as trust calibration [4] and several authors have shown that this can be achieved by sharing information about machine reliability with the operator (e.g. [5]).

Sharing situational measures of machine reliability fits in the domain of cooperative automation frameworks, which challenge designers and researchers to regard assistance functions or autonomous features as cooperative assistants, team players, or related terms [6–12]. These frameworks are based on requirements for cooperation and teamwork between humans identified in psychology and biology research.

Throughout the literature, key features for teamwork and cooperation (or team play, team cognition, partnership, joint work) that most authors seem to agree on, can be summarized:

- There is a basic agreement on a common goal.
- Members of the agreement (cooperation/team agents) possess a model of the current world view and state (e.g. mental state, situational competences) of other agents. This model can be updated through direct communication (e.g. stating why you are doing something or whether you are uncertain) or by inferring it from implicit cues (e.g. body language). The availability of decision justification information or state information of an agent contributes to *transparency.*
- Members can influence other agents in order to keep the goal/agreement intact and to align mismatching world views.

A characteristic of human-machine cooperation is that agents can share updates about current beliefs, decisions and uncertainties. When an agent receives indications of, for instance, a struggling team member, the informed agent may adjust its preparedness to rely on the struggling agent (i.e. recalibrate trust) and take more individual responsibility in the common task.

Human-machine interfaces that share this information on behalf of the machine are known as uncertainty displays, reliability displays or confidence displays.

For the current review, we collected examples from literature where reliability information was shared by the machine. The aim of this literature study was to obtain an overview of recent attempts, to identify knowledge gaps, and to extract guidelines that may be useful for the design of reliability displays.

METHODS

To obtain a collection of papers that contained implementations of reliability displays, we used Scopus and Google Scholar. Keywords related to reliability displays were combined with keywords relating to human-machine interaction studies ('operator', 'driver'). In Scopus, the following search query was used: ((“uncertainty display” OR “reliability display” OR “uncertainty communication” OR “reliability display” OR “confidence communication” OR “trust calibration”) AND (“operator” OR “driver” OR “robot”)). In Google Scholar, we combined the keys in front of the AND operator with words from behind the AND operator for individual searches, resulting in a total of 6x2 individual searches.

The resulting literature was screened based on the title and abstract, to confirm that one or more reliability displays were evaluated in the study. Only applications that provided real-time reliability information were considered. No constraint was placed on the publication type: since we wanted the most recent examples, conference and adjunct proceedings were also regarded. From the papers, relevant references were also taken into consideration. In the following section, we

review the communication methods that we found in these papers.

RESULTS

Overview

We found 16 papers that met the criteria. Seven papers were journal articles, six conference proceedings and two adjunct proceedings. In these papers, a total of 55 encoding variants were evaluated. Table 1 provides an overview.

A few types of studies were encountered. In eight studies, one reliability display was evaluated in a simulated environment. The main body of work (5) of these simulator studies consisted of driving scenarios. One was an aviation study, and another was from the military domain. In the simulated environments, the machines reliability was degraded, usually in parallel with changing weather conditions. In accordance, an uncertainty display became present or adjusted its intensity levels. Amongst other measures, we found that the effects of the display on measures of performance, safety and trust were most extensively studied.

Instead of studying one display in a simulated environment, several other studies [13–15] compared multiple displays or visual variables for their potential use as an uncertainty display. Participants were asked to order displays that presented varying reliability levels, allowing the researchers to investigate whether there was a consensus about the perceived reliability. This approach allowed for a quick evaluation of multiple encodings and gave a clear overview of ordering. A pitfall of this approach is that the participants do not experience the presented encodings in an environment where a task with shared responsibility needs to be executed. These studies were typically used as preliminary evaluation of multiple encoding variables, from which the preferred display was evaluated in an individual study.

Throughout the literature, authors spoke in terms of reliability, confidence or uncertainty. As most studies evaluated machines that became less reliable and thus needed to communicate a state of uncertainty, the latter of these three terms was the most popular choice of wording. In the following, we consider reliability displays, confidence displays and uncertainty displays as similar concepts and we may use them interchangeably.

Displaying machine uncertainty may have the goal to let the operator engage more in the shared task. Therefore, instead of expressing its own uncertainty, a machine could express to what extent it expects the driver to be engaged in the task. Noah et al. [14] investigated if displaying a more human-centered metric that they named 'required driving engagement' improved understanding. Twelve display designs that shared machine uncertainty were compared with displays that shared different levels of required driving engagement. It was found that uncertainty displays were generally better understood. We found that all of the currently selected studies provided information that reflected system-centered measures of uncertainty or reliability.

Methods of communicating reliability information

Most studies (14/16) offered reliability information to the visual modality. Except for one study where an auditory tone accompanied a flashing icon [16], we found no methods that employed audio. We found one example that made use of a haptic interface [17] and one application that used smell [18] to convey reliability information.

We further discuss these displays in the current section, structured by display type.

Qualitative displays

Icons Several studies used a binary indicator to communicate that the system was uncertain. One was a car simulator study by Beller et al. [19], who used an emoji-like icon showing a confused face reaching out with open palms to indicate system uncertainty. They showed that the uncertainty display increased the safety (time to collision) in situations where automation (level SAE 2)[1] failed. The data indicated improved situation awareness and better knowledge of automation weaknesses. Automation with the uncertainty symbol received higher trust ratings and increased acceptance.

Another binary indicator of uncertainty was designed by Louw and Merat [16], in the form of a flashing yellow icon of a steering wheel visible on the instrument cluster. The flashing icon was accompanied by a short tone, indicating that the vehicle was uncertain. The uncertainty visualization was not the main topic of the study, and no effects on safety or performance effects of the uncertainty displays were documented. The only finding by the authors regarding the uncertainty display was that the visual uncertainty display was not of much influence on the visual attention dispersion of the participants, and the authors suggested that the display did not require much effort for visually monitoring such displays. Unfortunately, the study did not provide evidence to support that the uncertainty display was not simply ignored.

Smell One of the few encountered displays that did not solely rely on the visual modality was one by Wintersberger et al. [20]. An olfactory interface was installed in a car simulator (SAE 2), that spread a lavender odor through the cabin for situations with high automation reliability and a lemon smell

[1] SAE international defined six levels of driving automation, from SAE Level 0 (no automation) to SAE Level 5 (full vehicle autonomy) [37].

for situations with low automation reliability. The olfactory display was combined with a visual icon on the instrument cluster, indicating high reliability with a green icon showing 'ACC' (Adaptive Cruise Control). Low system reliability was specified with an icon of a shoe that appeared to press a gas pedal or a brake. It was found that the addition of the olfactory component to the visual display decreased the intensity of brake pedal actuations and increased performance of a secondary detection-response task. Subjective reports indicated positive acceptance ratings of the olfactory display. Based on these results, the authors argued that the use of scents as encodings for system reliability is a promising method that deserves further attention.

Quantitative displays

Graphs and bars Instead of communicating a general state of vehicle uncertainty, more detailed representations of machine reliability can be found in displays with multiple levels or continuous measures reliability. In a flight simulator experiment, McGuirl and Sarter [5] accompanied an automated decision support system with a continuous line graph that indicated the system confidence for the past five minutes until now. The presence of the display resulted in better task performance and more appropriate responses to system errors. The authors concluded that this observation indicated more appropriately calibrated trust in the system and better human-machine team performance.

Helldin et al. [21] investigated the impact of visualizing assistance uncertainty on drivers' trust by displaying a visualization of assistance (SAE 2) competence in a driving simulation with varying weather conditions. The amount of machine confidence was displayed by means of seven empty bars that filled up as confidence increased, in a similar way to mobile phone status bars displaying signal quality. Drivers who were presented with reliability information took control of the car faster when needed, compared to the group without information about automation reliability. They also concluded that the reliability-info-group spent more time looking at non-driving related things, suggesting that they may be better able to perform tasks other than driving without compromising safety. Those presented with uncertainty feedback also reported less trust in the automated system. Helldin et al. concluded that this indicated a more proper trust calibration.

Kunze et al. [22] designed an anthropomorphic reliability display for a simulated SAE-level 3 automated vehicle. They made a visual display showing a peak from a heartbeat graph that lit up according to a simulated heartbeat frequency between 50 bpm (high reliability) 140 bpm (low reliability). The stylized graph seemed to have been inspired by medical heart rate monitors that can be found next to hospital beds of critical patients, usually characterized with a typical beeping tone (which was not audible in this study). In addition to the graph, a numeric value of the current machine heart rate was visible. The display resulted in safer takeovers of surrounding vehicles and in observable differences in gaze behavior in correspondence with different level of uncertainty. As a shortcoming of the display, the authors mention an increased workload and less accurate execution of secondary tasks, likely caused by the necessity to visually inspect the display regularly.

Further embodiments of quantitative reliability displays using graphs include a pie chart [23] and various proposals in a comparative study by Noah et al. [14].

Distortion, degradation, opacity Actively degrading or adding distortions to visual information can successfully convey different levels of uncertainty [23,24]. A similar encoding for reliability information is the manipulation of the opacity of the presented information, as applied in two studies [15,25]. Participants may prefer these types of encodings since the encodings are analogous to manifestations of uncertainty in real-world phenomena where uncertainty is caused by the absence, variability or ambiguity of available information [26]. For example, when road visibility conditions become worse due to fog, there is less information available to conclude where surrounding objects are. So, the natural degradation of signals can invoke a sense of uncertainty by the observer. Reflecting this in a data representation may therefore successfully communicate varying levels of system reliability.

Hue In a car simulator experiment where a car had autonomous capabilities (SAE 2), Faltaous et al. [27] encoded a continuous measure of uncertainty using an LED strip (30 LEDs, approx. 50 cm) capable of taking on a color from a full range of hue values between 0-100. This encoding did not turn out as successful method of communicating different ranges of assistance uncertainty. A problem that arose was that many participants did not perceive the display as a continuous range, but instead as three distinct states (red, yellow and green).

Another attempt to encode uncertainty information through color was made by Kunze et al. [28]. Also using a peripheral LED strip (77 LEDs, 50cm), they limited the hue range to the colors blue and red (cold-warm analogy). This encoding was successful in communicating differences between four distinct levels of uncertainty. Moreover, participants gave positive responses when asked whether they found the encoding logic for uncertainty communication. Similar conclusions were obtained by these authors in a different study, where subjects compared 11 different uncertainty visualizations for on-lane augmented projections [15]. Hue came out as the most preferred projection encoding when compared to 10 other visual encoding methods. The authors used a blue-purple-red color scheme, presumably similar to the color scheme that they used for their experiment with the LED strip.

Another study where reliability representations based on hue were correctly ordered by participants is by Noah et al. [13].

In this study, hue values were limited to five discrete levels, representing red, orange, yellow, light green and dark green.

The difference between the unsuccessful attempt by Faltaous et al. [27] and the successful attempts by Kunze et al. [15,28] and by Noah et al. [13] may be explained by data visualization guidelines [29,30] that advocate careful consideration when selecting hue ranges to represent continuous or multi-level information. Hue values that are physically ordered are not necessarily perceptually ordered; hue-based color maps can appear as if separated into bands of seemingly constant hue, with sharp transitions between the bands [29].

Vibrations In a comparative study of signal design parameters for a haptic car seat, Kunze et al. [17] evaluated if amplitude, position, movement and rhythm of vibrating signals could express different levels of vehicle uncertainty. The authors asked the participants whether they found that presented signals appeared to communicate an increase in uncertainty, or a decrease in uncertainty. Based on their results, the author argued that increases in vibration amplitude as well as rhythmic patterns consisting of long vibrations (3s) separated by short breaks of 0.5s communicated that uncertainty thresholds were reached, however, they were only able to communicate increases in uncertainty levels. The authors recommended not to use the proposed vibrotactile stimuli to display decreases in machine uncertainty.

Displaying reliability with reliable behavior

In a human-robot interaction study, Chen et al. [31] integrated a machine representation of human trust into the robot's decision making strategy. In a collaborative task, a human and machine had to clear a table of various objects. The human had the possibility to intervene and put objects away by himself, though without human interference a higher reward was given. Failure to clear the table led to different penalties, depending on the severity of the mistake: higher penalties were given for dropping wine glasses than for dropping fish cans. No penalty was given for dropping a plastic bottle.

By integrating a dynamic estimate of human trust into the robots decision making strategy, the researchers made the arm demonstrate competence by picking up the low-risk plastic bottles first. There were fewer human interventions for the wine glasses if the robot first demonstrated successful bottle removal, suggesting increased trust in automation. The authors argued that the inverse may also work: demonstrating failure or signs of incompetence may be used as a tool to lower the human's trust to more appropriate levels.

One could argue that a simple application of this relationship is already present in some commercially available vehicles. For example, Volvo's assistance functionality named Pilot Assist disables automated steering when it detects that a user is not interacting with the steering wheel for a prolonged amount of time [32], indicating an overreliance in the assistance functionality. The assistance functionality displays fallibility by letting the car drift sideways without intervening. This display of fallacious behavior forces the user to take back control, and potentially recalibrates trust in the assistance towards a more appropriate level.

Discussion

Studies with a focus on human-machine interaction with applications in the automotive, robotics, aviation and military domains have demonstrated the benefits of integrating automation with reliability displays. We created an overview based on the proposed communication methods in an attempt to identify knowledge gaps and extract guidelines to the design.

Modalities. We saw that most studies used visual encodings of reliability information. In comparison to other modalities, the visual modality may have the advantage that information can be transferred fast and with a high information density. However, a drawback is that the effectiveness of communication via this modality is dependent on the human's gaze direction. When there are other tasks at hand that require visual attention, like observing the road or engaging in non-driving tasks, drivers may neglect continuous visual displays [33]. To counter this, researchers have scarcely communicated reliability information using peripheral displays, haptics, or smell. We found no examples that primarily employed the auditory modality. We argue that the employment of the auditive, somatosensory and olfactory modalities to communicate reliability information deserves more attention.

Environmental modulations. To evaluate the effects of reliability communication methods, researchers have often created dynamic environmental conditions that altered the sensory performance of the machine agent. These conditions consisted for example of snow [21] or fog [16,19,22,27]. These conditions provide understandable causes of machine uncertainty to the human; snow and fog typically cause uncertainty to the human operator as well. However, causes of machine uncertainty that were less addressed in the studied literature are scenarios that unexpectedly cause machine uncertainty. Examples are the presence of light rain, which may affect a vehicle's observations obtained with Lidar [3], and according to the most recent Audi A8 manual [34], driving through a tunnel can affect radar observations. Except to the few who have read the manual before using the machine, these potential causes of uncertainty may be unknown. Exploring the effect of reliability displays when there is no congruency between the automation's and human's reliability may be an interesting topic for future research.

Trust assessments A returning topic in studies about reliability communication is trust, and in researchers' evaluations of reliability displays we often encountered

subjective assessments of trust such as Jian et al.'s scale [35], adjusted versions thereof or custom questionnaires.

Based on the outcomes of the trust assessments, some authors suggested that a machine that expressed uncertainty made the human trust the machine less and that this was a promising finding that indicated a more proper trust calibration [21]. Other authors stated that they found that trust was increased when machines were able to express uncertainty [19]. This second observation is often framed as a positive finding as increased trust can be an indicator of increased transparency [25]. This conclusion is understandable though it also shows that subjective trust measures should be interpreted with caution. Experimenters that find decreased levels of trust may be tempted to frame their finding as being supportive for trust calibration, and experimenters that find significant increases in trust may frame their finding as indicative of increased transparency. Decreases in trust should only be framed as desired effects in cases of overreliance and increases in trust are desired in cases of underreliance. To discriminate between either case, it is important to identify what the initial situation was in which the uncertainty was communicated: was the operator initially relying on the automation too much, or was the operator initially too skeptic or distrusting towards the automation's capabilities? Only when this initial attitude has been recognized, or forcibly set by the experimenters, one can judge whether proper trust calibration took place [4].

Another confusing factor about the measurement of trust for the evaluation of reliability displays lies in the fact that trust in automation can be increased when a participant is aware that the automation can make the statement that it is uncertain – a statement that by itself should, paradoxically, lower trust. Perhaps, a clearer distinction should be made to discriminate between overall trust in the automation, representing overall trustworthiness of the machine, and short-term trust, representing the willingness to situationally rely on the automation given the current circumstances. We have not found such distinctions in the studies encountered in this review, however in related research we found one attempt by Rajaonah et al. [36], who made separate questions about trust in the cooperation with automation, trust in self and trust in the automation when they evaluated a cooperative driver assistance feature. Unfortunately, the authors concluded from the answers to those questions that it was too difficult for participants to distinguish between these sub-types of trust. Future investigations related to this topic could focus on establishing frameworks and assessment techniques that comprise the notion of these different types of trust.

CONCLUSION

In this review, we collected research about the design and evaluation of machine reliability displays. We selected 16 papers from robotics, aviation, the military and automotive domains that discussed a total of 55 methods of encoding real-time measures of machine reliability. We have found the recruitment of visualizations most popular, though haptic feedback and the olfactory modality have also been employed. The majority of the studies have successfully demonstrated the benefits of integrating automated systems with reliability displays. Based on the work, we identified knowledge gaps and design guidelines that could be of help in the future design and evaluation of reliability displays.

Table 1: Overview of studies

key	year	1st author	n	modality	encoding variables	domain	publication type	encoding	dependent variables
[24]	2002	Finger	20	visual	1	ergonomics	Journal Article	distortion, vagueness	sorting task, ordering task, rating task
[5]	2006	McGuirl	30	visual	1	aviation	Journal Article	confidence line graph for past five minutes	pilot performance, compliance, accuracy estimate
[23]	2011	Neyedli	30	visual	2	military	Journal Article	pie chart, mesh (degradation)	Reliance, performance
[21]	2013	Helldin	59	visual	1	automotive	Conf Proceedings	display amount of bars machine competence bars	steering angle, brake, acceleration, look away times, time to takeover, trust (Jian modified)
[19]	2013	Beller	28	visual	1	automotive	Journal Article	uncertain face icon	trust, acceptance, interview
[13]	2016	Noah	36	visual	9	automotive	Adjunct Proceedings	compared 9	
[25]	2016	Mercado	30	visual	1	military	Journal Article	color opacity supported with textual explanation	performance, trust, workload
[14]	2017	Noah	21	visual	12	automotive	Conf Proceedings	compared 12	Sorting task
[16]	2017	Louw	60	visual, auditory	1	automotive	Journal Article	short tone, flashing yellow icon	Gaze distribution
[15]	2018	Kunze	46	visual	11	automotive	Conf Proceedings	compared 11	sorting task, response time
[17]	2018	Kunze	25	haptic	4	automotive	Adjunct Proceedings	static and dynamic vibromotor patterns on back	Subjective responses
[28]	2018	Kunze	25	visual	6	automotive	Conf Proceedings	led strip brightness, hue, position, size, movement, pulse	user responses.
[27]	2018	Faltaous	20	visual	1	automotive	Conf Proceedings	led strip color	trust, acceptance, perceived system reliability
[31]	2018	Chen	81	visual	1	robotics	Conf. Proceedings	fallacious behaviour	trust
[18]	2019	Wintersberger	25	olfactory, visual	2	automotive	Conf Proceedings	scent, icons	NDRT performance, trust (Jian), acceptance
[22]	2019	Kunze	34	visual	1	automotive	Journal Article	heartbeat animation combined with numerical display	trust, eye tracking, SA (SAGAT), Workload (NASA-TLX), performance of NDRT (visual search), MTTC, acceleration

REFERENCES

[1] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Hum. Factors* 39, 2 (1997), 230–253.

[2] N. Bagheri and G.A. Jamieson. 2005. The impact of context-related reliability on automation failure detection and scanning behaviour. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 212–217. DOI:https://doi.org/10.1109/icsmc.2004.1398299

[3] Sinan Hasirlioglu, Alexander Kamann, Igor Doric, and Thomas Brandmeier. 2016. Test methodology for rain influence on automotive surround sensors. *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC* (2016), 2242–2247. DOI:https://doi.org/10.1109/ITSC.2016.7795918

[4] John D Lee and Katrina A See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Hum. Factors J. Hum. Factors Ergon. Soc.* 46, 1 (January 2004), 50–80. DOI:https://doi.org/10.1518/hfes.46.1.50_30392

[5] John M McGuirl and Nadine B Sarter. 2006. Supporting Trust Calibration and the Effective Use of Decision Aids by Presenting Dynamic System Confidence Information. *Hum. Factors J. Hum. Factors Ergon. Soc.* 48, 4 (December 2006), 656–665. DOI:https://doi.org/10.1518/001872006779166334

[6] Joseph C R Licklider. 1960. Man-Computer Symbiosis. *IRE Trans. Hum. Factors Electron.* HFE-1, 1 (March 1960), 4–11. DOI:https://doi.org/10.1109/THFE2.1960.4503259

[7] Jean-Michel Hoc. 2001. Towards a cognitive approach to human–machine cooperation in dynamic situations. *Int. J. Hum. Comput. Stud.* 54, 4 (April 2001), 509–540. DOI:https://doi.org/10.1006/ijhc.2000.0454

[8] Klaus Christoffersen and David D Woods. 2002. How to make automated systems team players. In *Advances in Human Performance and Cognitive Engineering Research Volume 2*. Elsevier, 1–12. DOI:https://doi.org/10.1016/S1479-3601(02)02003-9

[9] Glen Klein, D.D. Woods, J.M. Bradshaw, R.R. Hoffman, and P.J. Feltovich. 2004. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intell. Syst.* 19, 06 (November 2004), 91–95. DOI:https://doi.org/10.1109/MIS.2004.74

[10] Patricia Bockelman Morrow and Stephen M Fiore. 2013. *Team Cognition: Coordination across Individuals and Machines*. Oxford University Press. DOI:https://doi.org/10.1093/oxfordhb/9780199757183.013.0012

[11] Guy Hoffman and Cynthia Breazeal. 2004. Collaboration in Human-Robot Teams. In *AIAA 1st Intelligent Systems Technical Conference*, 6434. DOI:https://doi.org/10.2514/6.2004-6434

[12] Matti Krüger, Christiane B Wiebel, and Heiko Wersing. 2017. From Tools Towards Cooperative Assistants. In *Proceedings of the 5th International Conference on Human Agent Interaction - HAI '17*, 287–294. DOI:https://doi.org/10.1145/3125739.3125753

[13] Brittany E. Noah, Thomas M. Gable, and Bruce N. Walker. 2016. Ordinal Magnitude Scaling for Automated Lane Keeping Displays. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - Automotive'UI 16*, 153–158. DOI:https://doi.org/10.1145/3004323.3004341

[14] Brittany E Noah, Thomas M Gable, Shao-Yu Chen, Shruti Singh, and Bruce N Walker. 2017. Development and Preliminary Evaluation of Reliability Displays for Automated Lane Keeping. In *Proceedings of the 9th International Conference on Automotive User*

Interfaces and Interactive Vehicular Applications - AutomotiveUI '17, 202–208. DOI:https://doi.org/10.1145/3122986.3123007

[15] Alexander Kunze, Stephen J. Summerskill, Russell Marshall, and Ashleigh J. Filtness. 2018. Augmented Reality Displays for Communicating Uncertainty Information in Automated Driving. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18*, 164–175. DOI:https://doi.org/10.1145/3239060.3239074

[16] Tyron Louw and Natasha Merat. 2017. Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transp. Res. Part C Emerg. Technol.* 76, (March 2017), 35–50. DOI:https://doi.org/10.1016/j.trc.2017.01.001

[17] Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. 2018. Preliminary Evaluation of Variables for Communicating Uncertainties Using a Haptic Seat. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18*, 154–158. DOI:https://doi.org/10.1145/3239092.3265959

[18] Philipp Wintersberger, Dmitrijs Dmitrenko, Clemens Schartmüller, Anna-Katharina Frison, Emanuela Maggioni, Marianna Obrist, and Andreas Riener. 2019. S(C)ENTINEL. In *Proceedings of the 24th International Conference on Intelligent User Interfaces - IUI '19*, 538–546. DOI:https://doi.org/10.1145/3301275.3302332

[19] Johannes Beller, Matthias Heesen, and Mark Vollrath. 2013. Improving the driver-automation interaction: An approach using automation uncertainty. *Hum. Factors* 55, 6 (2013), 1130–1141. DOI:https://doi.org/10.1177/0018720813482327

[20] Philipp Wintersberger, Dmitrijs Dmitrenko, Clemens Schartmüller, Anna Katharina Frison, Emanuela Maggioni, Marianna Obrist, and Andreas Riener. 2019. S(C)ENTINEL: monitoring automated vehicles with olfactory reliability displays. 538–546. DOI:https://doi.org/10.1145/3301275.3302332

[21] Tove Helldin, Goran Falkman, Maria Riveiro, and Staffan Davidsson. 2013. Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 210–217.

[22] Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. 2019. Automation transparency: implications of uncertainty communication for human-automation interaction and interfaces. *Ergonomics* 62, 3 (March 2019), 345–360. DOI:https://doi.org/10.1080/00140139.2018.1547842

[23] Heather F Neyedli, Justin G Hollands, and Greg A Jamieson. 2011. Beyond Identity: Incorporating System Reliability Information Into an Automated Combat Identification System. *Hum. Factors J. Hum. Factors Ergon. Soc.* 53, 4 (August 2011), 338–355. DOI:https://doi.org/10.1177/0018720811413767

[24] Richard Finger and Ann M Bisantz. 2002. Utilizing graphical formats to convey uncertainty in a decision-making task. *Theor. Issues Ergon. Sci.* 3, 1 (January 2002), 1–25. DOI:https://doi.org/10.1080/14639220110110324

[25] Joseph E. Mercado, Michael A. Rupp, Jessie Y. C. Chen, Michael J. Barnes, Daniel Barber, and Katelyn Procci. 2016. Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management. *Hum. Factors J. Hum. Factors Ergon. Soc.* 58, 3 (May 2016), 401–415. DOI:https://doi.org/10.1177/0018720815621206

[26] Georges-Pierre Bonneau, Hans-Christian Hege, Chris R Johnson, Manuel M Oliveira, Kristin Potter, Penny Rheingans, and Thomas Schultz. 2014. Overview and State-of-the-Art of Uncertainty Visualization. In *Scientific*

Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization, Charles D Hansen, Min Chen, Christopher R Johnson, Arie E Kaufman and Hans Hagen (eds.). Springer London, London, 3–27. DOI:https://doi.org/10.1007/978-1-4471-6497-5_1

[27] Sarah Faltaous, Martin Baumann, Stefan Schneegass, and Lewis L. Chuang. 2018. Design Guidelines for Reliability Communication in Autonomous Vehicles. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18*, 258–267. DOI:https://doi.org/10.1145/3239060.3239072

[28] Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. 2018. Evaluation of Variables for the Communication of Uncertainties Using Peripheral Awareness Displays. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18*, 147–153. DOI:https://doi.org/10.1145/3239092.3265958

[29] David Borland and Russell Taylor Ii. 2007. Rainbow Color Map (Still) Considered Harmful. *IEEE Comput. Graph. Appl.* 27, 2 (March 2007), 14–17. DOI:https://doi.org/10.1109/MCG.2007.323435

[30] B.E. Rogowitz and L.A. Treinish. 1998. Data visualization: the end of the rainbow. *IEEE Spectr.* 35, 12 (December 1998), 52–59. DOI:https://doi.org/10.1109/6.736450

[31] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2018. Planning with Trust for Human-Robot Collaboration. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, 307–315. DOI:https://doi.org/10.1145/3171221.3171264

[32] Volvo. 2018. Volvo V60 Owner's Manual. 315.

[33] Guy Cohen-Lazry, Avinoam Borowsky, and Tal Oron-Gilad. 2017. The Effects of Continuous Driving-Related Feedback on Drivers' Response to Automation Failures. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 61, 1 (September 2017), 1980–1984. DOI:https://doi.org/10.1177/1541931213601974

[34] Audi. 2018. A8 Owner's Manual. 96.

[35] Jiun-Yin Jian, Ann M. Bisantz, and Colin G. Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *Int. J. Cogn. Ergon.* 4, 1 (March 2000), 53–71. DOI:https://doi.org/10.1207/S15327566IJCE0401_04

[36] B. Rajaonah, F. Anceaux, and F. Vienne. 2006. Study of driver trust during cooperation with adaptive cruise control. *Trav. Hum.* 69, 2 (2006), 99. DOI:https://doi.org/10.3917/th.692.0099

[37] SAE International. 2016. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles.