Enhancing pedestrian crossing alerting systems

Hugo Boer, Max Bottemanne, Geert van den Broek, Stijn Mijnster and Wouter van der Wal

June 3, 2019

Abstract

More and more modern cars are fitted with alerting systems for crossing pedestrians. However, these systems act on the urgency of the situation (urgency-adjusted alerting systems) and produce warnings only when the situation becomes critical. In this research it was investigated whether an alerting system that acts on the awareness of the driver (awareness-adjusted alerting system) could have a positive impact on the response time of the driver. Additionally, it was investigated if such a system improves the acceptance of the driver. Using an experimental vehicle from the TU Delft (Toyota Prius), which is fitted with an eye tracking system, an awareness-adjusted alerting system was designed and tested by 21 test subjects on a closed test track in Delft. The eye tracker was used to determine if a driver was aware of a crossing pedestrian. In this experiment, the awareness-adjusted alerting system was compared to a baseline (no alerting at all) and an urgency-adjusted alerting system. The handling response time was improved by using an awareness-adjusted alerting system, whereas the gaze response time remained the same. The acceptance also did not show a significant difference.

Keywords – Eye-tracking, Human Behaviour, Alerting system, Pedestrians

1 Introduction

Pedestrians being killed by vehicles is still a worldwide problem. In 2018, the World Health Organisation (WHO) reported over 290 000 fatal collisions between pedestrians and cars [16]. To counter this problem, more and more cars are being fitted with pedestrian crossing alerting systems. These systems however, only give warnings for last-minute cases as warnings for each crossing pedestrian would become annoying and most of the times, redundant [19].

So what if the alerting system would give warnings based on the awareness of the driver instead of the urgency of the situation? Such a system would filter out the annoyingly redundant warnings, improving the communication between system and driver. This communication is important for the functionality of the system [19].

So how would such a system work? First, the system has to be able to determine whether the driver has spotted the pedestrian. Secondly, the system has to be able the determine the potential risk of a collision. The second condition is already met by most of the high tech cars today. To meet the first condition, eye tracking technologies, which are implemented in the TU Delft Toyota Prius, could be used.

There have been multiple studies on how to warn a driver of potential hazards [4, 15, 17, 19, 20], and on the driver's observation performance [18, 26, 27]. Research on the combination of the two however is still scarce. The research about how to warn a driver of potential hazards could be used in designing an awareness-adjusted crossing alerting system. But there are still some challenges in defining whether a driver has actually seen an object [26].

This brings us to the following research question: does an awareness-adjusted alerting system (by means of eye tracking), opposed to a fixed urgency alerting system, improve the driver's response time concerning crossing pedestrians along with the driver's acceptance towards the alerting system?

In this study, a real vehicle driving experiment was carried out to determine whether an awareness-adjusted alerting system improves the driver's acceptance and response time. This is tested in three different situations : (1) a baseline test without any alerting, (2) a test with an alerting system that acts on urgency and (3) a test with an alerting system that acts on the driver's awareness. This paper uses the key terms and parameters as defined by [21]. It also applies standards for equipment and procedures defined by [22].

In the following chapter, an explanation of the theory used to set up the method is given. The study is done by bachelor students of Mechanical Engineering at the Technical University of Delft as part of their Bachelor thesis.

2 Theory

2.1 Alerting systems in cars

A widely discussed topic related to alerting systems is the communication towards the driver. The interaction with a collision warning system can help faster and more efficient responses, but at the same time can require a monitoring task and evaluation process that may lead to automation complacency [19]. In particular, reliable warnings can decrease brake reaction time when compared to misleading warnings, which slow visual search for hazard detection. A lack of driving experience slows down the overall response, while unexpected failure of the system leads to unintentional blindness. The effectiveness of the warning also depends on the modality (auditory vs. visual) and specificity (low vs. high) of the alerting system [20]. Specific visual warnings tend to have the most positive effects on reaction time and collision risks, while specific auditory warnings were considered to be the most annoying.

Theories on alerts in cars are also investigated for Take-Over Requests (TORs) applied in automated vehicles. At automation level 3 or higher [13], a driver is allowed to conduct Non-Driving Related Tasks (NDRTs). Performing these NDRTs lead to loss of situation awareness. This becomes dangerous if the automated system reaches its boundaries, in which case a TOR has to be given. According to [4] there are three main ways to communicate this alert to the driver, namely through auditory, visual and vibrational messages. Of these three, the auditory messages through beeps, bell sounds and horns were considered to be the most annoying.

The frequency of the alerts plays a vital role in the effectiveness of the alerting system. If the system gives alerts too frequently, the driver is annoyed by the system. If the system only gives alerts right before a crash, the driver may not respond quickly enough [17]. The frequency of the alerts additionally has an impact on the amount of false alarms. If the frequency increases, then so does the amount of false alarms. This can cause a decrease in reliability of the system. A lower reliability causes a decrease in the appropriate driving responses [7]. If the frequency of the alerts decreases, the system will be less able to detect objects and could even miss events. A good balance in alert frequency is therefore necessary.

When designing an alerting system, it is important to investigate the visual behaviour of the driver. Mainly the glance location and the glance eccentricity are essential for compiling the visual behaviour [15]. Furthermore, three sensory cues (longitudinal deceleration, looming and brake lights) were found to be relevant for capturing the attention of the driver and increasing the glances to the forward path, influencing the visual behaviour.

2.2 Visual

The visual field can be divided into three fields, all at an angle about the point of fixation [26]: (1) the central visual field occupying about $\pm 4^\circ$ eccentricity of visual angle about the point of fixation, (2) the macular visual field located between 5° and 9° eccentricity and (3) the peripheral visual field located beyond 9° eccentricity. Ball and Owsley [3] distinguishes the peripheral visual field even further with a Useful Field Of View (UFOV), where any information that falls within the UFOV is processed in a single fixation and any information that falls outside of this region is not processed. The UFOV captures everything until 15°-20° eccentricity. According to [6], the visual field is categorized as the central visual field of up to about 30° eccentricity and the peripheral visual field beyond 30° eccentricity. The peripheral visual field provides awareness of larger targets and information on moving targets whereas the central visual field provides information on most targets. As pedestrians fall under the scope of larger targets, there is still no clear answer that these visual fields affect the detection of these pedestrians.

According to [23], there is an increasing threshold for the minimum speed of an object in these visual fields in order to detect them. This partly answers the above question, as slowly moving objects are detected less, and consequently slower in larger angles of eccentricity. However, the distinction between reaction times in various visual fields is not mentioned. Furthermore, [23] states that there is a minimum reaction time of 200 milliseconds to detect objects.

[3] also states that aging results in loss of the UFOV and increases the error rate in various degrees of eccentricity. What is frequently mentioned is that the visual fields and reaction times vary a lot between different people. It is hard to pin an exact threshold of reaction time for visual fields due to the lack of literature on this matter.

2.3 Eye tracking

There are several ways to detect the gaze of a driver. Each eye tracking system has its benefits and weaknesses. However, to determine the gaze of a driver, a system has to have three essential components: (1) robust facial feature tracking, (2) head pose and gaze estimation and (3) 3-D geometric reasoning [25]. The quality of the product varies a lot and is often related to the production costs of the system [10]. In [11], the quality of five different systems is compared, where the Smart Eye Pro has the best percentage of usable data (100%).

2.4 Galvanic Skin Response (GSR)

The galvanic skin response, also known as electrodermal activity or EDA, refers to the changes in sweat gland activity. These are reflective of the intensity of the human emotional state, otherwise known as emotional arousal [8, 9]. This activity of sweat glands is triggered by postganglionic sudomotor fibers in the human skin [5]. The skin conductance (SC) is measured by two electrodes that require skin contact in order to produce a reliable signal [2]. By applying a low constant voltage, the SC can be measured non-invasively [5].

The time series of SC can be characterized by a slowly varying tonic activity (i.e. skin conductance level, SCL) and a fast varying phasic activity (i.e. skin conductance response, SCR) [5]. The standard peak detection method (trough-to-peak) defines the SCR amplitude as the difference of the SC values at its peak and at the preceding trough.

[1] introduced a decomposition method by means of deconvolution. The method is based on the assumption that sudomotor nerve activity originally shows peaks (sudomotor bursts) with short time constants which trigger SCRs exhibiting larger time constants. The deconvolution of SC data with an appropriate impulse response function (IRF, also called transfer function) is intended to reverse this transformation. The IRF represents the basic SCR shape that would result from a unit impulse. A Bateman function (i.e., a bi-exponential function) was found to represent an adequate IRF in this deconvolution procedure.

3 Method

3.1 Specifications

The vehicle is equipped with a Smart Eye Pro eye tracker. So the choice of system was already made. But according to [11], the Smart Eye Pro system is the best eye tracker in their comparison, with 100% of usable data. The position and movement of the vehicle is obtained by using a differential GPS, wheel encoders and an inertial measurement unit. With visual odometry from the stereo vision cameras on the front of the vehicle, the system computes the coordinates of surrounding objects and pedestrians relative to its own coordinate system. After calibrating the eye-tracker on the eyes of the participant, the exact angle was measured between the direction of the gaze relative to the direction of the car.

The auditory alerting system was implemented in the vehicle by installing two speakers in the front of the car. One speaker is placed on the left side and one at the right side of the participant.

The participants were fitted with a push button around their index finger to measure the handling response time of a participant. Pushing the button created a time stamp which was used to calculate duration times.

3.2 Code

A python script running real time during the experiment, operating at both 10Hz and 60Hz, detected objects as “pedestrian” by a SSD algorithm at 10Hz. All pedestrians outside a 30 meter scope were filtered. Pedestrians within the 30 meter scope got a classification “not seen”.

The system identified if the pedestrian was in an early stage of potential danger for the vehicle. This was done using 3 preconditions. Firstly, the pedestrian had to be inside an 18 meter scope. With an average vehicle speed of 10-15 km/h this gave the participant an average of 5 seconds to reach the pedestrian. Secondly, the value of the angle between the car and the pedestrian had to be between $-50°$ and $50°$. This excluded pedestrians without the intention to cross the street, but was mainly for filtering false alarms. The system sometimes identified a pedestrian that went from $\pm-80°$ to $\pm70°$ in one iteration. This was definitely not a real pedestrian and restricting the angle filtered those false alarms.

As section *2.2* stated, the minimum reaction time, independent of the visual fields, was 200 ms. The operating speed of the eye tracker was 60Hz resulting in 12 iterations before participants could react. This was tested and concluded that the system operated too slowly. The number of iterations was lowered to 4 iterations with the knowledge that participants processed the appearance of pedestrians faster than they reacted on it. However, the system was tested and reviewed as too strict. A threshold of 8 iterations, and therefore a processing time of 133 ms, was eventually granted before the system could classify the pedestrian as “potential danger”.

A counter that iterates using the gaze angle (the angle between the gaze and the pedestrian operating at 60Hz) determined if the classification changed to “seen”. This was done using two visual fields, the UFOV and the central visual field (as explained in *2.2*). Since the UFOV processes everything within a single glance, whereas the central visual field takes longer, the counter iterated differently for both gaze angles. This was to ensure that the participant looked directly at the pedestrian. Using only the peripheral visual field was not enough. If the gaze angle was inside 18°, the counter iterated with factor 2. If the gaze angle was between 18° and 30°, the counter iterated with factor 1. The same threshold of “potential danger” was appointed to the classification of “seen”, namely 8 iterations.

3.3 Measurements

The system was tested on two main categories: (1) the response time of the participant to crossing pedestrians and (2) the acceptance of the participant towards the system. Both categories were expressed in two measurements. For the response time, these measurements were the gaze response time and the handling response time, which will be explained in the upcoming two sections. To test the acceptance towards the system, the galvanic skin response of the participant was measured and a post-experiment 'Van Der Laan' questionnaire was filled in by the participant [24].

3.3.1 Gaze Response Time

The gaze response time was measured as the difference between two timestamps. When the pedestrian was detected by the system, the system created a starting time stamp. When the pedestrian was classified as “seen”, the system produced a stopping timestamp. Subtracting both timestamps resulted in the corresponding gaze response time.

3.3.2 Handling Response Time

The handling response time was also measured as the difference between two timestamps. The starting time stamp of the handling response time was the same as the starting time stamp of the gaze response time. When the participant pushed the button, the system produced a stopping timestamp. Subtracting both timestamps resulted in the corresponding handling response time.

3.3.3 Galvanic Skin Response

The GSR was measured using the "Seeed 101020052 Grove" [12]. This sensor is supplied with two elasticated conductive pads. These are slit over the middle and ring finger's proximal phalanx. The sensor has an analogue output (the serial port reading (SPR)) which can be read by a Raspberry Pi. This output results from the voltage drop, caused by skin conductance, over a resistance the value was capped between zero (representing no skin resistance) and 512 (representing infinite skin resistance). To determine the skin conductance (SC), the following formula was used [12]:

$$\mathrm{SC} = \frac{(512 - \mathrm{SPR}) * 100}{(1024 + 2 * \mathrm{SPR})} \qquad \in \mu S \tag{1}$$

To analyze the SC, "Ledalab" was used [14]. This Matlab-based software can perform a continuous decomposition analysis (CDA), which performs a decomposition of SC data into continuous signals of phasic and tonic activity. This method takes advantage from retrieving the signal characteristics of the underlying sudomotor nerve activity (SNA) [5, 14]. The starting timestamp of pedestrian detection was labelled as "Event" in Ledalab. Concerning these timestamps, a time interval of zero to three seconds was investigated. Within these intervals the following measurements were computed: response latency of the first significant SCR (response of SC above threshold of 0,01 μS), maximum value phasic activity and mean tonic activity.

3.3.4 Questionnaire

After the experiment, the participants were asked to fill in a questionnaire about the experiment, which included an usefulness/acceptance questionnaire created by Van Der Laan [24]. In this 'Van Der Laan' questionnaire, the participants rated the systems with 5 questions on usefulness and 4 on acceptance. The answers were scaled from -3 to +3 and the positive and negative side of the scores were alternated. This resulted in three personal usefulness and acceptance scores (one per system) for each participant.

Besides the Van Der Laan test, it was questioned if the explanation of the experiment was clear, if the pedestrians were visible at the start of a run and if the participant was comfortable driving the vehicle. There was one multiple choice question which asked the personal preference of the participant. Afterwards the participants were asked about their thoughts of how each system worked.

3.4 Experiment setup

Prior to the experiment, the participant was briefed about all necessary information to perform the experiment. During the experiment, the participant had to drive over the test track with non driving related tasks (NDRTs). Three different systems were compared to each other, called system 0, system A and system B, explained in *section 3.4.3*. All systems were tested three times in a row resulting in a total of 9 test runs per person. After the experiment, the participant had to fill in the questionnaire.

3.4.1 Briefing

The participant was provided with all information needed to participate. They were told that pedestrians would appear arbitrary and that the non-driving related task had to be performed. Furthermore, they were asked to press the button if a pedestrian was recognized and to not use the brakes unless there was actual danger. Only a handful of information was intentionally omitted. The difference between system A and system B together with the true purpose of the eye tracker were omitted to ensure an unbiased result. The participant was specifically asked to pay attention to the difference of both systems A and B, to drive between 10 and 15 km/h and to drive like he or she would on a public road. The amount of crossing pedestrians was told to vary between 2 and 3.

3.4.2 Test track

The test track was located on a closed area behind the faculty of Mechanical, Maritime and Materials Engineering of Delft University of Technology, which is depicted in *figure 1*. The length of the test track was roughly 120 meters and eight obstacles in the form of garbage containers were placed in sets of two (one on the left and one on the right) alongside the road. In each run, the pedestrians appeared from behind three of the four sets of obstacles.

The participant was encouraged to look on either the left or the right side of the road due to the placement of a NDRT on either the left or the right container. This NDRT was either reading a word or solving an easy math problem. The size of writing was calibrated so the participant could read it from a distance of about 15 meters. The pedestrians either appeared on the same side of the NDRT or on the opposite side.

To ensure the safety of the pedestrians and the driver, a chalk line was drawn on the road one meter next to each container. After the pedestrians appeared from behind the container, they walked to the line but not any further. The driver was told to drive between the lines of the left and right container to avoid potential collisions.

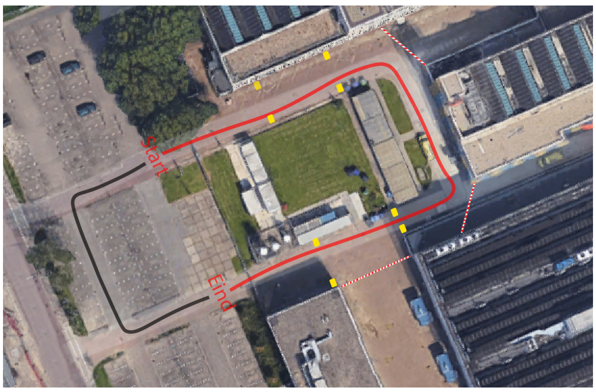

Figure 1: *Overview of the test track*

3.4.3 Three systems

As explained before, three different systems were compared to each other: system 0, system A and system B. All three systems were able to classify pedestrians as "potential danger" and "seen". System 0 was the baseline, giving no alarms even when the pedestrian was classified as "potential danger" and "not seen". System A always gave an alarm for pedestrians classified as "potential danger", independent of the "seen" classification. System B gave an alarm only for pedestrians classified as "potential danger" and "not seen".

3.4.4 Participants

The participants of the experiment were 21 students who were doing there masters at the Delft University of Technology. Every participant was in the possession of a driver's license for at least 4 years, had driven more than 2500 kilometres in the last year and was experienced with an automatic transmission. The group consisted of 15 male and 6 female participants between the age of 21 and 28 (M=23.8, SD=1.7).

3.5 False alarms

During the experiment, false alarms occurred. There was a distinction in false alarms between false positives and false negatives. False positives occurred when an alarm was given, where it was not needed. During the experiment, the participants were told by the co-driver when these type of false alarms were given. A false negative meant that an alarm was not given when the system actually had to give one.

3.5.1 False positives

False positives in this experimental setup occurred because of four reasons. Firstly, other pedestrians that were not part of the experiment interfered. Although the test track was a closed area, it was still an open area for employees, resulting in possible false positives. Secondly, weather conditions resulted in false positives. Sunny days caused overexposure and rainy days resulted in raindrops on the windscreen and therefore windscreen wipers appeared. These conditions obstructed the camera view resulting in extra pedestrians seen that did not exist. Thirdly, objects or specific combinations of objects were seen as a pedestrian, also resulting in false positives. Lastly, existing pedestrians were detected multiple times.

3.5.2 False negatives

False negatives in this experimental setup occurred because of two reasons. Firstly, the system did not recognize the pedestrian as a pedestrian or did not recognize the pedestrian at all. Secondly, the pedestrian was classified as "seen" while the participant did not see the pedestrian, due to eye tracker malfunctioning.

3.5.3 Predictive values

The false alarms were expressed in a negative predictive value (NPV) and a positive predictive value (PPV). The required parameters for the NPV could not be determined. Especially when the second type of false negative had to be determined, it became very difficult. It was not possible to determine with the bare eye whether a participant had actually seen the pedestrian.

The required parameters for the PPV could be extracted from the result. The PPV could be calculated through:

$$PPV = \frac{\#\text{true positives}}{\#\text{true positives} + \#\text{false positives}}$$

3.6 Expected results

The expected results were based on the literature study at the beginning of this project. When compared to an urgency alerting system, an awareness-adjusted alerting system could result in a decreased annoyance level because unnecessary alarms would be filtered out [17, 19]. Furthermore, the alarm effectiveness could be enhanced when the communication between system and driver is improved [17]. Two hypotheses were stated:

- The driver's response time on crossing pedestrians will decrease when using an alerting system that is awareness-adjusted
- The driver's acceptance towards an alerting system will be improved when it is awareness-adjusted.

4 Results

The results are divided into the two parameters where the system is tested on; the response time of the participant and the acceptance of the participant towards the system.

4.1 Response time

In total, there should be 567 timers to measure the gaze response time (GRT) and handling response time (HRT) of the system. However, due to faults in the system or pedestrians appearing too late, some data was lost. In the case of the HRT it could also be possible that the participant forgot to press the button. In *table 1* the remaining timers are displayed. This means that 32 gaze response times and 33 handling response times were lost.

Timer	system 0	system A	system B	**Total**
GRT	182	181	172	**535**
HRT	182	180	172	**534**

Table 1: *Sample sizes for GRT and HRT sorted per system and in total*

It is also useful to look at the gaze response times in the cases where potential danger was detected. As in real life, these are the situations where an alerting system could have an impact. The "dangerous situations" are defined as the situations where the participant is not aware of a pedestrian appearing from behind the container (in which case systems A and B would have given an alarm). When filtering on only "dangerous situations", the sample sizes in *table 2* are obtained. The sample size of the gaze response time and handling response time is the same.

Timer	system 0	system A	system B	**Total**
GRT	46	49	40	**135**
HRT	46	49	40	**135**

Table 2: *Sample sizes in dangerous situations for the GRT and the HRT sorted per system and in total*

The data is splitted into "same side", "other side" and "both sides". "Same side" means that the pedestrian appeared from the same side as where the NDRT was displayed. "Other side" means that the pedestrian appeared from the opposite side of where the NDRT was displayed. "Both sides" is the combined data of "same side" and "other side".

4.1.1 Gaze response time

The result of the average GRT per system is plotted with standard deviation in *figure 2*. An independent-samples t-test is conducted to compare system 0 and system A. There is a significant difference in the scores for system 0 (M=0.201, SD=0.230) and system A (M=0.284, SD=0.454), conditions; t(361)=2.134, p = 0.05. These results suggest that system A has an effect on the gaze response time of the participant. Specifically, these results suggest that with the use of an awareness adjusted alerting system, the gaze response time is decreased. The difference between system 0 and system B (M=0.241, SD=0.337) and between system A and system B, respectively under the conditions; t(352)=1.253, p = 0.30 and t(351)=1.008, p = 0.40, does not have a small enough p value and is therefore inconclusive. Furthermore, the average gaze response time is faster for "same side" than "other side". This would mean that the experiment setup was successful.

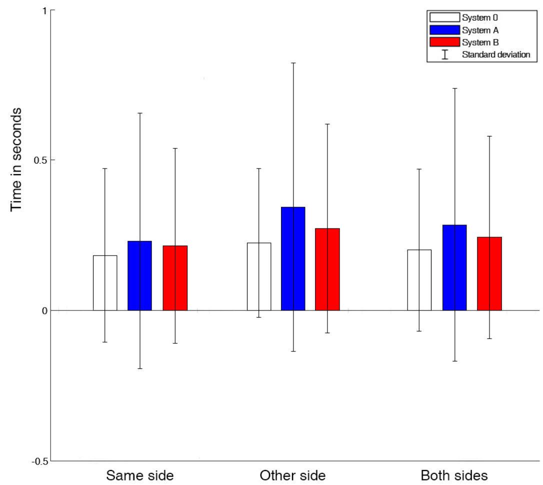

Figure 2: *Gaze response time of each system at all situations*

When comparing the three systems in dangerous situations, the GRTs (*figure 3*) are quite close. Again an independent-samples t-test is conducted to compare system 0, system A and system B. The difference in the scores for system 0 (M=0.449, SD=0.340) and system A (M=0.506, SD=0.603), conditions; t(93)=0.559, system 0 and system B (M=0.486, SD=0.432), conditions; t(84)=0.339 and system A and system B, conditions; t(87)=0.187 all have a p value higher than 0.50. This confirms that no possible conclusion can be drawn.

Looking at both *figure 2* and *figure 3*, one can conclude that the overall GRT is considerably slower in dangerous situations. The average GRTs for both situations are displayed in *table 3*. The result of this comparison is not a surprise, but a confirmation of what should be measured in these cases.

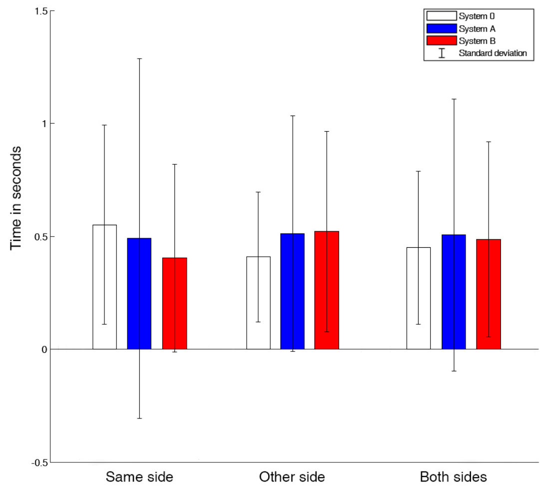

Figure 3: *Gaze response time of each system in dangerous situations*

Data	System 0 [s]	System A [s]	System B [s]
All situations	0.20056 (SD=0.26962)	0.28417 (SD=0.45420)	0.24110 (SD=0.33696)
Dangerous situations	0.44889 (SD=0.33922)	0.50554 (SD=0.60315)	0.48619 (SD=0.43228)

Table 3: *Average GRTs for both sides (all data and filtered on danger)*

4.1.2 Handling response time

The results for the handling response times are shown in *figure 4*. An independent-samples t-test is conducted to compare system 0 and system B. There is a significant difference in scores for system 0 (M=0.161, SD=0.547) and system B (M=0.039, SD=0.460), conditions; t(360)=2.260, $p = 0.05$. These results suggest that system B has an effect on the handling response time of the participant. Specifically, these results suggest that with the use of an awareness adjusted alerting system, the handling response time is decreased. There is also a difference in scores for system A (M=0.132, SD=0.504) and system B, conditions; t(350)=1.797, $p = 0.10$. These results suggest that there is a 10% probability of the data being the outcome that there is no difference between system A and B. Comparing system 0 and system A results in a p value higher than 0.50 and is therefore inconclusive.

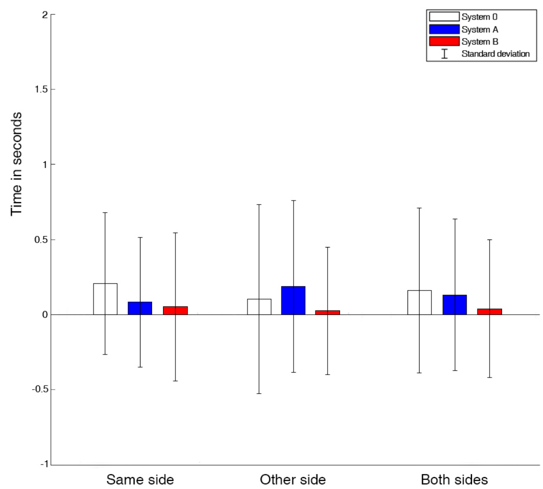

Figure 4: *Handling response time of each system at all situations*

On average over all systems, it took the participants 0.11 seconds to press the button after observing the pedestrian, keeping in mind that the system needs processing time before recognizing the pedestrian.

What could be discussed is that the participants might have gotten used to the experiment and consequently responded increasingly faster as the experiment lasted. This is in comparison with system 0, as system A and system B were alternated to remove the effects of bias.

In *figure 5* the results of the handling response times are shown for the dangerous situations. By looking at the 'both sides' graph on the right, which gives the best view of reality, it can be noticed that the handling response time is improved by using alerting systems in comparison with system 0. Again an independent-samples t-test is conducted to compare system 0, system A and system B. There is a significant difference in the scores for system 0 (M=0.432, SD=0.658) and system A (M=0.363, SD=0.596), conditions; t(93)=0.534, $p > 0.50$, system 0 and system B (M=0.154, SD=0.663), conditions; t(84)=1.950 $p = 0.10$ and system A and system B, conditions; t(87)=1.568, $p = 0.20$. These results suggest that system B has an effect on the handling response time of the participant. Specifically, these results suggest that an awareness adjusted alerting system improves the handling response time. With less confidence it can be said that there is a difference between system A and system B. No conclusion can be determined for the difference between system 0 and system A.

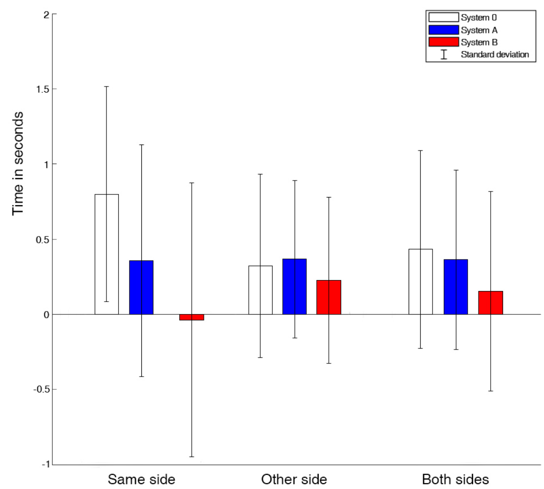

Figure 5: *Handling response time of each system in dangerous situations*

In *table 4* the average handling response times for all data and the dangerous situations are given. The same conclusion can be drawn as in the case of the GRT; the participant reacts slower when there is danger in comparison to all other situations. This is a confirmation that the classification "danger" is appropriate in this situation.

Data	System 0 [s]	System A [s]	System B [s]
All situations	0.16083 (SD=0.54738)	0.13151 (SD=0.50359)	0.03902 (SD=0.46009)
Dangerous situations	0.43227 (SD=0.65825)	0.36346 (SD=0.59638)	0.15382 (SD=0.66339)

Table 4: *Average HRTs for both sides (all data and filtered on danger)*

4.2 Acceptance

The acceptance of the systems is firstly tested by measuring the Galvanic Skin Response (GSR) of the participant. The GSR is a measurement of emotional arousal, which can be related to acceptance. In total there are 9 datasets per participant (3 systems x 3 runs). This brings the total to 189. The results of the GSR measurements are described in section 4.2.1.

The second measurement tool for acceptance is the questionnaire, which was filled in by the participants after the experiment without knowing how each system worked. The results are described in section 4.2.2.

4.2.1 Galvanic Skin Response

Figure 6 shows the raw GSR data of a single run. *Figure 7* shows the same GSR data after CDA. The red lines indicate an event (a moment when a pedestrian appeared). The grey surface shows the tonic activity and the blue surface corresponds to the phasic activity. The sharp peaks could be an artifact from the pedestrian pressing the button. We have tested this and found no conclusive reason as to what caused the peaks.

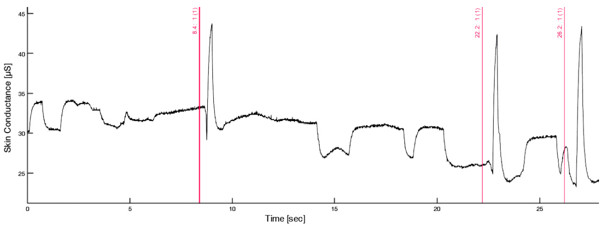

Figure 6: *Raw data in Ledalab before CDA.*

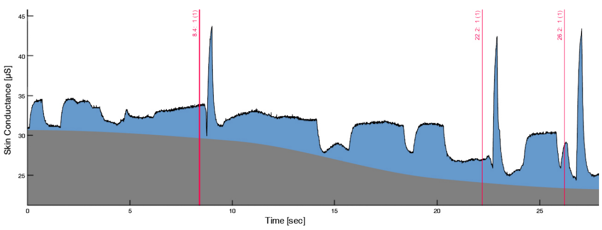

Figure 7: *Processed data in Ledalab after CDA.*

All raw data from the GSR sensor was analyzed through CDA after which all results were exported. The average values with standard deviations are depicted in *table 5*.

Results	Latency $[s]$	PhasicMax $[\mu S]$	Tonic $[\mu S]$
System 0	0.544 (SD=0.562)	21.74 (SD=60.74)	8.60 (SD=9.31)
System A	0.593 (SD=0.532)	26.26 (SD=83.78)	10.76 (SD=10.83)
System B	0.589 (SD=0.572)	12.22 (SD=20.38)	9.84 (SD=11.17)

Table 5: *GSR results*

These results show that there is only a small difference in latency between the three systems. An independent-samples t-test is conducted to compare system 0, system A and system B. There is a difference in the scores for system 0 (M=0.386, SD=0.562) and system A (M=0.513, SD=0.532), conditions; t(311)=2.048, p = 0.05, system 0 and system B (M=0.409, SD=0.572), conditions; t(307)=0.355, p > 0.50 and system A and system B, conditions; t(308)=1.655, p = 0.10. The differences are very small. This delay could be due to human responsiveness. The maximum phasic activity has more varied results in which system B has the fewest high values. The scores from the independent-samples t-test are for system 0 (M=8.397, SD=60.743) and system A (M=7.619, SD=83.776), conditions; t(311)=0.094, for system 0 and system B (M=6.934, SD=20.375), conditions; t(307)=0.283 and for system A and system B, conditions; t(308)=0.098 where the value of p is higher than 0.50 in every situation and therefore inconclusive. There is a difference in the scores for the tonic activity for system 0 (M=3.987, SD=9.312) and system A (M=5.864, SD=10.832), conditions; t(311)=1.644, p = 0.10, system 0 and system B (M=4.002, SD=11.169), conditions; t(307)=0.355, p > 0.50 and system A and system B, conditions; t(308)=1.490, p = 0.20. These results suggest that there is respectively 10%, more than 50% and 20% probability that there is no difference between the three systems tested.

4.2.2 Questionnaire

Most of the sanity checks in the questionnaire were meant to validate the experimental setup after each run. *Figure 8* displays the average distance, relative to the car, at which the NDRT was completed.

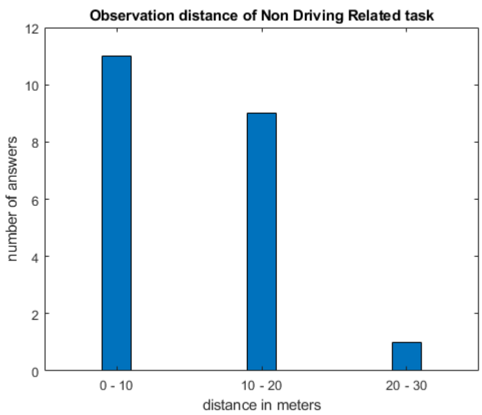

Figure 8: *The distance where the non driving related task was completed*

The result of the simple preference question was quite close. As displayed in *figure 9*, there was a slight advantage for system B (the awareness adjusted system). Furthermore, There were only 3 persons that understood the difference between system A and system B, and each of these 3 persons chose system B as their favourite system.

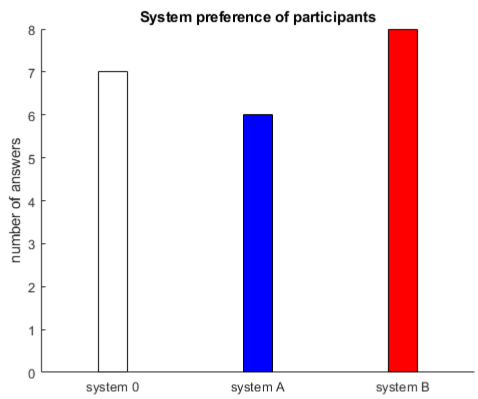

Figure 9: *System preference of the participants*

The Van Der Laan test resulted in an average score on acceptance and usefulness per participant. As explained in *section 3*, the system is tested on a scale from -3 to 3 with 4 questions on acceptance and 5 questions on usefulness. In *figure 10* a boxplot of usefulness scores is displayed, where system A has the highest score. An independent-samples t-test is conducted to compare system 0 and system A. There is a significant difference in the scores for system 0 (M=-0.276, SD=1.379) and system A (M=0.571, SD=1.132), conditions; t(40)=2.177, p = 0.05. These results suggest that system A has an effect on the usefulness. Specifically, these results suggest that with the use of a full alerting system, the system is experienced as more useful with a 5% probability that there is no difference. There are also differences in the scores for system 0 and system B (M=0.067, SD=1.199), conditions; t(40)=0.860, p = 0.20 and for system A and system B, conditions; t(40)=1.403, p =0.20. The p value in both situations is 0.20 and therefore inconclusive.

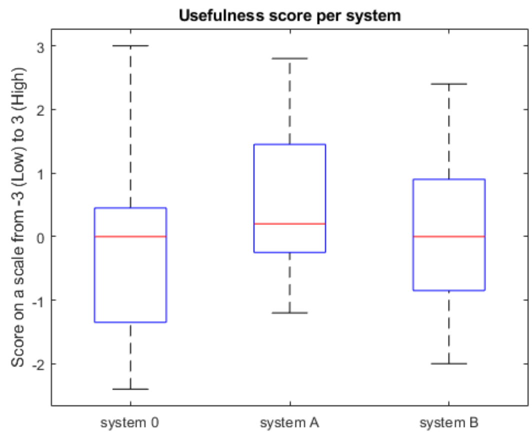

Figure 10: *Boxplot of the usefulness score of the three systems*

The acceptance of each system is also displayed in the form of a boxplot in figure *11*. In contrast to the usefulness, this plot shows a bigger difference. System 0 is the best accepted system, followed by system B and system A. This is in line with what is mentioned in *section 2* and [7, 17].

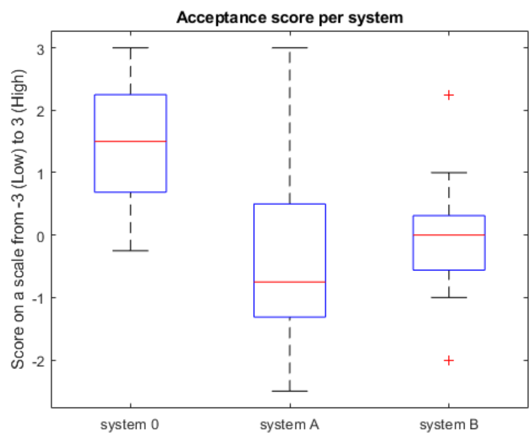

Figure 11: *Boxplot of the acceptance score of the three systems*

An independent-samples t-test is conducted to compare system 0 and system A. There is a significant difference in the scores for system 0 (M=1.476, SD=1.018) and system A (M=-0.333, SD=1.302), conditions; t(40)=5.016, p = 0.001. There is also a significant difference in the scores for system 0 and system B (M=-0.036, SD=0.863), conditions; t(40)=5.189, p = 0.001. These two results suggest that system A and system B have an effect on the acceptance of the system. Specifically, these results suggest that an alerting system is less accepted by the participants. The comparison between system A and system B with the conditions; t(40)=0.873, p = 0.20 is inconclusive with the fact that there is 20% probability that there is no difference between the two systems.

When looking at the usefulness and acceptance scores of the participants that successfully deducted the difference between system A and system B, the result is different, as depicted in *table 6* and *table 7*. The usefulness of system B is rated higher with an average score of 1.8. The acceptance had an average score of 1.3.

However, the sample size of this group is very small (3 persons). This sample size is too small to draw any conclusions from the result. Furthermore, this result is most certainly biased as the data is filtered with participants that have a correct understanding of the system.

Person	system 0	system A	system B
1	0.2	0	1.4
5	-2.4	1.6	1.6
18	-0.8	-0.2	2.4
Average	**-1**	**0.5**	**1.8**

Table 6: *Usefulness score of the people with a correct system understanding*

Person	system 0	system A	system B
1	1.5	-1	1
5	0.75	-1.5	0.75
18	2.25	-2	2.25
Average	**1.5**	**-1.5**	**1.33**

Table 7: *Acceptance score of the people with a correct system understanding*

4.3 False alarms

In *table 8*, the amount of false and true positives are given along with the corresponding positive predictive values (PPVs). The resulting PPVs clearly indicate that the system produced a considerable amount of false alarms. Sometimes these false positives were copies of the real pedestrian, which resulted in two signals in quick succession. False alarms due to random events resulted in alarms given on random moments. During rain, the raindrops and windscreen wipers obstructed the camera view resulting in falsely detected pedestrians. During sunny days, the reflection of the sun through buildings caused obstructions resulting in falsely detected pedestrians. Additionally the system falsely classified two umbrellas and a scooter parked on the side of the test track as pedestrians.

System	True positives	False positives	PPV
System 0	N.A.	N.A.	N.A.
System A	187	114	62.13%
System B	41	44	48.24 %
Total	**228**	**158**	**59.07%**

Table 8: *The amount of false and true positives per system, and the resulting PPV*

5 Discussion

The aim of this study was to examine the effects of an awareness-adjusted alerting system on the response time of the driver concerning crossing pedestrians and the acceptance of the driver towards the system. In the experiment, three different systems (system 0, A and B, explained in *section 3.4.3*) were tested on 21 participants. The response time was determined through the gaze response time (GRT) and handling response time (HRT). For the acceptance, the galvanic skin response was measured and the participants filled in a Van Der Laan questionnaire [24].

5.1 Response time

The comparison between the different systems on GRT did not result in a clear difference between the three systems. There was only a significant difference between system 0 and system A. In this case system 0 was significantly faster than system A. This is a surprising result, as one would usually assume that system B is faster than system 0, especially in the cases that the driver has not seen the pedestrian.

The HRT showed a more significant difference. System B was significantly faster than both other systems when comparing all the data (“potential danger” or not). In situations of “potential danger”, system B was significantly faster than system 0. The comparison between system B and system A was inconclusive when looking at “potential danger”.

Overall, system B only improves the handling response time of the driver. There were no clear improvements in the gaze response time.

5.2 Acceptance

The GSR did not provide any conclusive data about the stress level of a driver towards the three different systems. This was mainly caused by the large standard deviations.

When looking at the results of the Van Der Laan questionnaire, it can be concluded that system A was experienced as more useful than system 0. However, system 0 was significantly more accepted than system A and system B. All other possible comparisons were inconclusive.

Interestingly however is the acceptance and usefulness score of the people who successfully deducted the functioning of system B. They rated system B higher on both usefulness and acceptance. Even though these people are biased in formulating their opinion, it is an interesting result because real life drivers are also aware of the capabilities of their vehicle.

Overall, both measurements were too inconclusive to determine what the effect on the acceptance of the three systems were. System 0 was the best accepted system, looking at the questionnaire, but this was affected by the lack of knowledge of the participants. Almost every participant, bare three, thought that system B was not correctly tuned or not sensitive enough. Most of these participants wanted to alter their answer after hearing how system B worked.

5.3 Limitations and shortcomings

The experiment falls short on the amount of false positives, mainly because of the weather conditions, but also due to system malfunctions, detecting scooters or umbrellas as pedestrians or creating two pedestrians from one “real” pedestrian. When looking at the positive predictive values (PPV) in *table 8*, both system A and B are performing poorly. With an average PPV of less than 60%, the system does not reach the desired quality to use in real life. There are some clear results in especially the HRT of all three systems, but to be fully sure of the result, the detection system has to be improved as there are a lot of false positives due to the detection system.

Another shortcoming is the delay of the system to identify an object as pedestrian, which in extreme cases could be 300 milliseconds. Therefore, a driver occasionally recognized a pedestrian and responded to the threat more quickly than the system. This is not particularly bad, as the pedestrian can still be classified as “seen” before there is “potential danger”. Only in the case where the participant already spotted the pedestrian and was looking away, after which the system yet spotted the pedestrian, this resulted in a false alarm. This however did not occur often and therefore did not have a substantial impact on the experiments.

In *section 3.2* it is explained which visual fields have been used to determine whether a driver has or has not seen a pedestrian. The gaze angle beyond 30° was left out as a visual field where a driver would (partly) see a pedestrian or his/her movement. From literature and our experience however, this part of the visual field is not as unusable as initially defined. Some of the drivers were capable of observing moving objects beyond the 30°. However, this also did not occur often and it therefore did not have a substantial impact on the experiments.

5.4 Recommendations

When looking at the results of the questionnaire in *section 4.2.2* and especially to those who successfully deducted the difference between the systems, one could think that the acceptance and usefulness look completely different when the difference between the systems is explained prior to the experiment. When the difference is told, bias is created, but when a car is bought, the driver also knows what systems are present in the vehicle. So one could conduct a test where participants are told about the functioning of each system prior to the experiment.

The detection of “pedestrians” could be improved. As the minimum reaction time of a driver concerning pedestrians was 200 milliseconds, as stated in *section 2.2*, the latency should at least be reduced to 100 milliseconds. This ensures that the system is within the 200 milliseconds window for checking “potential danger” with pedestrians. In this experiment, the cameras detecting pedestrians operated at 10Hz. Upgrading this to 60Hz could shorten the latency.

Something that could be reduced is the amount of false alarms. Reducing the amount of false alarms improves the PPV, resulting in a more reliable system. A PPV of over 90% removes all uncertainties about the quality of the system. To reduce the amount of false alarms, a better pedestrian detection system has to be implemented.

To interpret the results of the GSR, only one analyzing application was used (Ledalab, [14]). This is something that could be explored for further research.

6 Conclusions

In *section 3.6* the hypotheses are stated as follows:

- The driver's response time on crossing pedestrians will decrease when using an alerting system that is awareness-adjusted
- The driver's acceptance towards an alerting system will be improved when it is awareness-adjusted.

Following the results (*section 4*) of the experiment (described in *section 3*), the hypotheses can be rejected. Although the handling response time is significantly improved, there is no clear indication that there is a difference in gaze response time. Furthermore, when evaluating the acceptance of the system, the confidence in the difference between the systems is too low to draw any conclusions.

With an awareness adjusted alerting system, the distance where alarms are given could be increased. As in current situations, alarms are only provided in high urgency situations to reduce the amount of redundant and therefore annoying warnings. When using an ideal awareness-adjusted alerting system, the alarms are never redundant and could therefore be given when needed, instead of last-minute cases.

7 Acknowledgement

First of all, we would like to thank the 21 participants of the test for their time and cooperation to make the results of the experiments as useful as possible.

Furthermore, we would like to thank everyone at the Vehicle Engineering group for their help and for their flexibility in using the Toyota Prius.

Finally, we would like to thank our supervisor Jork Stapel. Without Jork we could not have finished our Bachelor thesis project in the way we have now. Jork invested a lot of time in this project and very frequently gave helpful feedback, for which we are deeply grateful.

References

[1] Alexander, David M et al. "Separating individual skin conductance responses in a short interstimulus-interval paradigm". In: *Journal of neuroscience methods* 146.1 (2005), pp. 116–123.

[2] Bakker, Jorn, Pechenizkiy, Mykola, and Sidorova, Natalia. "What's your current stress level? Detection of stress patterns from GSR sensor data". In: *2011 IEEE 11th international conference on data mining workshops*. IEEE. 2011, pp. 573–580.

[3] Ball, Karlene K et al. "Age and visual search: Expanding the useful field of view". In: *JOSA A* 5.12 (1988), pp. 2210–2219.

[4] Bazilinskyy, Pavlo et al. "Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays". In: *Transportation research part F: traffic psychology and behaviour* 56 (2018), pp. 82–98.

[5] Benedek, Mathias and Kaernbach, Christian. "A continuous measure of phasic electrodermal activity". In: *Journal of neuroscience methods* 190.1 (2010), pp. 80–91.

[6] Bhise, Vivek D. *Ergonomics in the automotive design process*. CRC Press, 2016, p. 68.

[7] Bliss, James P and Acton, Sarah A. "Alarm mistrust in automobiles: how collision alarm reliability affects driving". In: *Applied ergonomics* 34.6 (2003), pp. 499–509.

[8] Boucsein, Wolfram. *Electrodermal activity*. Springer Science & Business Media, 2012.

[9] Critchley, Hugo D. "Electrodermal responses: what happens in the brain". In: *The Neuroscientist* 8.2 (2002), pp. 132–142.

[10] Dong, Yanchao et al. "Driver inattention monitoring system for intelligent vehicles: A review". In: *IEEE transactions on intelligent transportation systems* 12.2 (2010), pp. 596–614.

[11] Funke, Gregory et al. "Which eye tracker is right for your research? Performance evaluation of several cost variant eye trackers". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 60. 1. SAGE Publications Sage CA: Los Angeles, CA. 2016, pp. 1240–1244.

[12] *Grove GSR Sensor Technical Question (Resistance Equation)*. URL: https://community.seeedstudio.com/Grove-GSR-Sensor-Technical-Question-(Resistance-Equation)-t-13059.html.

[13] International, SAE. *SAE J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles RATIONALE*. SAE International, 2016.

[14] *Ledalab*. URL: http://ledalab.de/.

[15] Morando, Alberto, Victor, Trent, and Dozza, Marco. "Drivers anticipate lead-vehicle conflicts during automated longitudinal control: Sensory cues capture driver attention and promote appropriate and timely responses". In: *Accident Analysis & Prevention* 97 (2016), pp. 206–219.

[16] Organization, World Health et al. *Global status report on road safety 2018*. World Health Organization, 2018.

[17] Parasuraman, Raja, Hancock, Peter A, and Olofinboba, O. "Alarm effectiveness in driver-centred collision-warning systems". In: *Ergonomics* 40.3 (1997), pp. 390–399.

[18] Roth, Markus, Flohr, Fabian, and Gavrila, Dariu M. "Driver and pedestrian awareness-based collision risk analysis". In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2016, pp. 454–459.

[19] Ruscio, Daniele, Ciceri, Maria Rita, and Biassoni, Federica. "How does a collision warning system shape driver's brake response time? The influence of expectancy and automation complacency on real-life emergency braking". In: *Accident Analysis & Prevention* 77 (2015), pp. 72–81.

[20] Schwarz, Felix and Fastenmeier, Wolfgang. "Augmented reality warnings in vehicles: Effects of modality and specificity on effectiveness". In: *Accident Analysis & Prevention* 101 (2017), pp. 55–66.

[21] Standardization, International Organization for. "Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems, Part 1: Definitions and parameters". In: *ISO 15007* (2014).

[22] Standardization, International Organization for. "Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems, Part 2: Equipment and procedures". In: *ISO 15007* (2014).

[23] Tynan, Paul D and Sekuler, Robert. "Motion processing in peripheral vision: Reaction time and perceived velocity". In: *Vision research* 22.1 (1982), pp. 61–68.

[24] Van Der Laan, Jinke D, Heino, Adriaan, and De Waard, Dick. "A simple procedure for the assessment of acceptance of advanced transport telematics". In: *Transportation Research Part C: Emerging Technologies* 5.1 (1997), pp. 1–10.

[25] Vicente, Francisco et al. "Driver gaze tracking and eyes off the road detection system". In: *IEEE Transactions on Intelligent Transportation Systems* 16.4 (2015), pp. 2014–2027.

[26] Wolfe, Benjamin et al. "More than the useful field: considering peripheral vision in driving". In: *Applied Ergonomics* 65 (2017), pp. 316–325.

[27] Yang, Yucheng et al. "An HMI Concept to Improve Driver's Visual Behaviour and Situation Awareness in Automated Vehicle". In: *International Conference on Intelligent Transportation Systems (ITSC*. 2018, pp. 650–655.