VTI rapport 639A
Published 2009

www.vti.se/publications

Results of a field study on a driver distraction warning system

Katja Kircher
Albert Kircher
Christer Ahlström

vti
FINDING A BETTER WAY

Publisher:

SE-581 95 Linköping Sweden

Publication:
639A

Published:	**Project code:**	**Dnr:**
2009	40658	2006/0226-26

Project:
IVSS Distraction and Drowsiness

Author:
Katja Kircher, Albert Kircher and Christer Ahlström

Sponsor:
Saab, within IVSS programme

Title:
Results of a field study on a driver distraction warning system

Abstract (background, aim, method, result) max 200 words:

The main goal of the distraction and drowsiness field study was to evaluate a system for detecting driver distraction and drowsiness. This report focuses on the results of the study, indicating how a distraction warning system influenced glance behaviour.

A vehicle was instrumented with an automatic eye tracker and other sensors. Seven participants drove the vehicle during one month each. During the first ten days a baseline was collected. Afterwards the warnings were activated, which involved that the drivers received a vibration in the seat when the algorithm determined that they had looked away from the forward roadway for a too long time.

The main finding was that the drivers' gaze behaviour was not influenced much by the distraction warnings. The drivers received distraction warnings at about the same frequency during the treatment and the baseline phase. Performance indicators like "percent road centre" and others did not change from baseline to treatment phase. The average percentage of very long glances decreased slightly in the treatment phase, suggesting that the warning had an effect on the more extreme glance behaviour. There are also indications that the system helped prevent further extended glances away from the road immediately after a warning was issued.

Keywords:
distraction, field study, eye tracking, drowsiness

ISSN:	**Language:**	**No. of pages:**
0347-6030	English	96

Utgivare:	Publikation:		
	639A		
581 95 Linköping	Utgivningsår: 2009	Projektnummer: 40658	Dnr: 2006/0226-26
	Projektnamn: IVSS Distraktion och sömnighet		

Författare:
Katja Kircher, Albert Kircher och Christer Ahlström

Uppdragsgivare:
Saab, inom IVSS-programmet

Titel:
Resultat av en studie av distraktionsvarningssystems påverkan på förarnas blickbeteende

Referat (bakgrund, syfte, metod, resultat) max 200 ord:

Huvudmålet med distraktions- och sömnighetsstudien var att utvärdera ett system för att detektera förardistraktion och sömnighet. Denna rapport fokuserar på studiens resultat, huruvida ett distraktionsvarningssystem påverkade förarnas blickbeteende.

En försöksbil utrustades med en automatisk "eye tracker" och andra sensorer. Sju deltagare körde fordonet under en månad var. Under de första tio dagarna samlades data om vanligt körbeteende utan varningssystem (baseline). Sedan aktiverades varningarna, vilket innebar att förarna fick en varning i form av en vibration i förarsätet när algoritmen bedömde att de hade tittat bort från vägen för länge.

Huvudresultatet var att förarnas blickbeteende inte förändrades mycket på grund av distraktionsvarningssystemet. Förarna fick distraktionsvarningar med ungefär samma frekvens när varningarna var på som under baseline-fasen. Prestationsindikatorer som ”percent road centre” och liknande ändrades inte heller mellan baseline-fasen och när varningarna var på. Den genomsnittliga procentandelen av långa blickar bort sjönk något när varningarna var på, vilket innebär att varningarna hade en effekt på det mer extrema blickbeteendet. Det finns även tecken på att varningarna förhindrade fortsatta längre blickar bort från vägen precis efter en varningsaktivering.

Nyckelord:
distraktion, fältstudie, eye tracking, sömnighet

ISSN:	Språk:	Antal sidor:
0347-6030	Engelska	96

Preface

This report is the last in a series of three on an extended field study on driver distraction financed by the IVSS programme. The first report (Kircher, K., 2007) covers the theoretical background of driver distraction with special focus on glance behaviour and traffic safety. The second report (Kircher, K., Kircher, A. & Claezon, F., 2009) describes the instrumented vehicle and the experimental setup in detail. The present report contains the results of the study.

Christer Ahlström, VTI, contributed with a large part of the analyses done in Matlab, and with valuable suggestions and discussion points. Albert Kircher, VTI, provided the chapter on self-reported results and was an important partner for discussions. Without him and Fredrich Claezon (then Saab, now Scania) it would not have been possible to conduct the study.

Special thanks go to Arne Nåbo at Saab for taking us on board, and to the IVSS programme for financing this study. We would also like to thank SmartEye AB for support with the eye tracker. Last but not least we thank all participants who were willing to share their data with us, and who made this study possible.

Linköping March 2009

Katja Kircher

Cover: photos.com and Katja Kircher, VTI

Quality review

Review seminar was carried out on 18th of December 2008 where Astrid Linder reviewed and commented on the report. Katja Kircher has made alterations to the final manuscript of the report. The research director of the project manager Jan Andersson examined and approved the report for publication on 10th of March 2009.

Kvalitetsgranskning

Granskningsseminarium genomfört 2008-12-18 där Astrid Linder var lektör. Katja Kircher har genomfört justeringar av slutligt rapportmanus 2009-03-10. Projektledarens närmaste chef, Jan Andersson, har därefter granskat och godkänt publikationen för publicering 2009-03-10.

Table of Contents

Results of a field study on a driver distraction warning system

by Katja Kircher, Albert Kircher and Christer Ahlström
VTI (Swedish National Road and Transport Research Institute)
SE-581 95 Linköping Sweden

Summary

An extended field study on driver distraction and drowsiness was conducted in an instrumented vehicle, which was equipped with an eye tracker and other sensors. This report focuses on the findings concerning driver distraction. Seven participants used the car just as their own car during a period of one month each. During the baseline phase, which comprised of the first ten days of the trial the distraction warnings were deactivated. During the treatment phase, consisting of the remaining 20 days, the warnings were activated, meaning that the driver received a vibration in the seat whenever an algorithm called AttenD determined that the driver was distracted from the driving task. The participants' subjective opinion about the warning systems was assessed with the help of three questionnaires.

The method is promising for driver distraction research, as it investigates naturalistic behaviour in a naturalistic setting. The employed eye tracker held up to the expectations, even though it is recommendable for future research to use more than two cameras. With the current setup, there was a tendency that tracking was lost just when driver distraction occurred. A robust data acquisition system is a requirement.

The main finding was that the drivers' gaze behaviour was not influenced much by the distraction warnings. The drivers received distraction warnings at about the same frequency during the treatment phase as they would have during the baseline phase. This indicates that they did not avoid the warnings. Performance indicators like "percent road centre" and the newly developed percentage of glances within the "field relevant for driving" did not change from baseline to treatment phase. The standard deviation of gazes did not change, either. The average percentage of very long glances decreased slightly in the treatment phase, suggesting that the warning had an effect on the more extreme glance behaviour. There are also indications that the system helped prevent further extended glances away from the road immediately after a warning was issued.

The results from the questionnaire indicate that the drivers were satisfied with AttenD. Their expectations had been positive, and they indicated no disappointment. The drivers stated that they trusted the system, that the warnings were not experienced as disturbing, and that the system made them more aware of what they did while driving. Some drivers reported using their cell phones less while driving as a consequence of the warnings.

The analyses presented here are of a rather general nature, and more detailed analyses could provide new insights and a more differentiated picture of the usefulness of the driver distraction warning system. It is also important to investigate whether AttenD influenced driving behaviour like speed choice or steering variables.

A general problem with driver distraction research is the absence of a ground truth, which could be used as a benchmark, against which distraction detection algorithms could be compared and evaluated.

Resultat av en studie av distraktionsvarningssystems påverkan på förarnas blickbeteende

av Katja Kircher, Albert Kircher och Christer Ahlström
VTI
581 95 Linköping

Sammanfattning

Detta är den tredje och sista delen i en rapportserie om förardistraktion. Serien är baserad på en längre fältstudie om trötthet och distraktion i trafiken och denna rapport fokuserar alltså på de resultat som rör distraktion. Sju deltagare använde en instrumenterad bil som utrustats med ett distraktionsvarningssystem kallat AttenD under en månad vardera. De första tio dagarna utgjorde den så kallade "baseline"-fasen, och då var distraktionsvarningssystemet avstängt. Under de resterande 20 dagarna var varningssystemet på, vilket innebar att förarsätet började vibrera när distraktionsalgoritmen ansåg att föraren var distraherad från köruppgiften. Försökspersonernas subjektiva uppfattning av varningssystemet undersöktes med hjälp av tre enkäter.

Den instrumenterade bilen var bland annat utrustad med en datalogger och ett eye tracking-system. Den eye tracker som användes levde upp till förväntningarna även om studien gjordes under tuffa fältförhållanden. Vi rekommenderar dock att fler än två kameror används i framtida studier. Med utrustningen som användes här kunde vi se en tendens att trackingen tappades just när förarna blev distraherade. En annan viktig lärdom från projektet är att ett stabilt dataloggsystem är ett krav för en fältstudie av denna typ.

Huvudresultatet från studien var att förarnas blickbeteende inte förändrades mycket på grund av distraktionsvarningssystemet. Förarna fick distraktionsvarningar med ungefär samma frekvens när varningarna var på som de skulle ha fått under baseline-fasen. Detta visar att de inte undvek varningarna. Prestationsindikatorer som ”percent road centre” och procentandelen av blickar inom ”field relevant for driving”, som togs fram i denna studie, ändrades inte heller mellan baseline-fasen och när varningarna var på. Standardavvikelsen av blickriktningen förändrades inte heller. Den genomsnittliga procentandelen av långa blickar (mer än 2 sekunder) bort från vägen sjönk något när varningarna var på, vilket innebär att varningarna hade en effekt på det mer extrema blickbeteendet. Det finns även tecken på att varningarna förhindrade fortsatta längre blickar bort från vägen precis efter en varningsaktivering.

Resultaten från enkäterna visar att förarna var nöjda med AttenD. Deras förväntningar på systemet var positiva och de visade ingen besvikelse efter försöksperioden. Förarna uppgav att de litade på systemet, att varningarna inte upplevdes som störande och att de kände att systemet hjälpte dem att bli mer medvetna om vad de höll på med medan de körde. Några förare rapporterade att de använde sina mobiltelefoner mindre under körningen på grund av varningarna.

Analyserna som presenteras här är på en övergripande nivå och mer detaljerade analyser skulle kunna leda till nya insikter och en mer differentierad bild av användbarheten av distraktionsvarningssystemet. Det är också viktigt att undersöka om AttenD påverkade körbeteenden såsom hastighetsval och styrrelaterade parametrar.

Ett generellt problem med förardistraktionsforskningen är att det saknas en etablerad sanning om vad distraktion egentligen är som skulle kunna användas som mått med vilket distraktionsdetektionsalgoritmer kan jämföras och evalueras.

1 Introduction and background

This report is the third in a series of three, focusing on the results of the IVSS Inattention and Drowsiness project. The first report in the series covers a literature review on driver distraction with focus on eye gaze behaviour and its relation to driving behaviour and traffic safety (Kircher, K., 2007), the second report gives a detailed description of the method used, including a chapter on “lessons learnt” (Kircher, K., Kircher, A. & Claezon, F., 2009). In order to be able to make most of the present report it is highly recommended to read through the second report, which gives a detailed overview of the method employed in this study. In this report only parts of the method that are immediately necessary for understanding the results are taken up, for details the reader is referred to the other two reports in the series.

The goal of the study was to evaluate a real time distraction mitigation system called AttenD and a drowsiness mitigation system in a long-term field test in a natural setting. Due to the reason that the eye tracker used in the study had not been subjected to this kind of field test before, and that the method used was quite new for all partners involved, another goal of the study was to evaluate both the equipment and the method itself.

Simulator studies had shown that it was difficult to attain “true distraction” in an artificial setting (e.g. Almén, 2003; Karlsson, 2005). Distraction mitigation researchers, who performed a series of experiments in different simulators recommend using a field test for further evaluation of distraction mitigation systems (e.g. Donmez, Ng Boyle & Lee, 2006, 2007; Donmez, Ng Boyle, Lee & McGehee, 2006; Zhang & Smith, 2004). These results, together with the maturation of remote eye tracking systems, which can now be operative for a long time without experimenter intervention, led to the decision to perform a distraction mitigation test in the field, using the general methodological setup of a field operational test (FOT), but on a smaller scale than common for this type of test.

Seven drivers used an instrumented Saab 9-3 in their daily lives for about a month each. During the baseline phase, which consisted of the first approximately 10 days, the driver was not given any feedback on his behaviour; the car “behaved” just like a normal car. The driving behaviour, including warnings that would have been given, was logged.

During the treatment phase, which lasted for the remaining approximately 20 days, logging continued, and warnings for inattention and drowsiness were not only logged but also given to the driver. The drivers’ subjective opinion about the warning systems and their expectations and experiences were obtained via a set of questionnaires and by interviewing the drivers.

In order to help interpreting the results a short description of the inattention detection algorithm is given here. A more detailed version can be found in (Kircher, K., Kircher, A. & Claezon, F., 2009).

The algorithm AttenD is based on a defined visual vehicle model, which divides the vehicle into different zones like the windshield, the speedometer, the mirrors, the dashboard, etc., and on the time the driver spends glancing at those zones. A buffer is decremented over time when the driver looks away from the “field relevant for driving” (FRD), which consists of the intersection between a circle of a visual angle of 90° and the vehicle windows, excluding the area of the mirrors. When the driver’s glance is inside the FRD, the buffer is incremented again, until a maximum value of 2 s is reached. Special latencies for decrementing are built in for the mirrors and the speedometer, recognising the need to check mirrors and speedometer for traffic safety reasons. There is a delay of 0.1 s for increasing the buffer again after having been decreased, in order to compensate for focal adaptation and an “adaptation of the mind” to the road scene and away from the secondary task that had been attended. When the buffer reaches zero the driver is considered to be distracted, and when certain further conditions were met, which are described in detail in Kircher et al. (2009), a warning was given to the driver.

The described rules apply as long as gaze direction tracking data of a quality of at least 0.25 is present. Otherwise the algorithm relies on head/nose direction tracking with comparable rules, or, if head/nose tracking is not available either, on a simple decision rule for no-tracking cases, trying to keep false alarms at an acceptable level without missing crucial warnings.

The study was conducted in order to find out whether the distraction warning system AttenD would influence behaviour such that drivers would look away from the FRD less with the system active, which was supposed to increase traffic safety. Several hypotheses are related to this research question. The most immediate effect would be that AttenD cuts off glances away from the forward roadway, making drivers look up faster than they would have without a warning. This question is taken up in Chapter 8. Another effect could be that drivers try to avoid warnings, meaning that the frequency of driver distraction, and thus the number of warnings, decreases with the distraction warning system activated. This issue is addressed in Chapter 4. Furthermore, the amount of very long glances, that are considered to be especially dangerous, might be reduced with a distraction warning system. Glance duration is considered in Chapter 5. Last but not least, a distraction warning system could generally have an effect on how drivers distribute their glances in the environment. They might focus their attention more on the road centre and other areas relevant for driving. This is investigated in Chapter 6.

The report is structured such that for each chapter a short introduction of the topic in question is presented, and after the presentation of the results a short discussion of the topic is included. A more general and comprehensive discussion can be found in the end of the report.

2 Data preparation

For each driver both video logs and text logs were available. More detailed descriptions of the log data can be found in Kircher et al. (Kircher, K., Kircher, A. & Claezon, F., 2009). The participants are numbered from 10 to 16, in the chronological order in which they participated in the study.

The text logs came from seven modules. One came from the GPS and included position and speed. The remaining six modules shared a common time stamp and logged a host of data from the sensors installed in the vehicle, and from the CAN. They had the following format in common: The filename consisted of a code for the vehicle, the driver, the module, the day of the year and the time of day when the file was created. All those six log files belonging to the same trip therefore carried the same filename save for the module ID. Examples are presented in Table 1. This nomenclature allowed fast and unambiguous identification of files and trips. The name structure was used when programming analysis code.

Table 1 Examples for filenames from different modules.

Filenames from Module 8050 (IDP)	Filenames from Module 8100 (CAN-bus)
`da-1-11-8050_071114_233631.log`	`da-1-11-8100_071114_233631.log`
`da-1-11-8050_071115_081952.log`	`da-1-11-8100_071115_081952.log`
`da-1-11-8050_071115_083747.log`	`da-1-11-8100_071115_083747.log`

The following six modules existed:

8010: Warning control module (3 variables)

8020: Raw SmartEye data module (21 variables)

8030: Drowsiness module (13 variables)

8050: Inattention Detection Programme (IDP) module (25 variables)

8070: In/Out module (5 variables)

8100: CAN bus data (25 variables)

Therefore, excluding the two video channels and the GPS log, 92 variables were logged continuously, including two time stamps that were equal across all six parallel modules. A more detailed description of exactly which variables were logged in which module can be found in Kircher et al. (Kircher, K., Kircher, A. & Claezon, F., 2009) describing the method of the study.

All six modules shared the time stamp, but the rows for each module were written at slightly different times. This was due to the fact that the modules consisted of separate programmes, which ran in parallel. The operating system Windows determined via a scheduler in which order the processor executed those different programmes. This switching between processes was very fast and therefore execution appeared to be parallel. There was no predetermined order in which the processes were executed, therefore it was not determined either in which order between processes the data rows were written to memory. In general the time delay until the measured data was committed to memory was very short, from around 0.025 s to 0.1 s, depending on the module. For analyses that drew on data from different modules the row with the nearest time stamp was considered to be the matching information.

For analysis the data had to be reduced. Several different reduction methods were used. One was trip-based, for more general analysis of driver behaviour, such as length of trip, distribution of speed across the trip and the like. Glance frequency analyses were either based on weighted trip-based data or on absolute counts of glances. Another reduction method was event-based, where each occurrence of inattention as determined by the AttenD algorithm counted as event. This reduction method was mainly used for the reaction time related behaviour and warning frequency.

Matlab code was written for the extraction of relevant data. For some analyses further processing was done in SPSS.

Due to the small number of participants and the large variation between participants with respect to a number of variables, not many inferential tests were done. Where they were done, they were usually kept within participants.

The treatment phase was about twice as long as the baseline phase for most participants. Therefore for most analyses comparisons were not only made between baseline and treatment phase, but the treatment phase was split into three periods of similar duration, as was also done by LeBlanc et al. (2006). This can be seen as a simplified method to account for some time series effects. The exact number of days during which the participants could use the car varied between drivers, therefore the time periods could have slightly different lengths across participants. These periods are called “weeks” or “phases” in the analyses, with week 1 corresponding the baseline condition, week 2 or phase 2 being the first week in the treatment condition, week 3 or phase 3 being the second week in the treatment condition and week 4 or phase 4 being the last week in the treatment condition. An overview of the number of days per week per participant can be found in Table 2. Participant 16 had a somewhat shorter driving period than the other participants, due to approaching summer holidays. For Participant 14 week 2 was interrupted, because the project sponsors required the car to be present at a press event.

For all analyses except those on trip statistics the first two days of the baseline phase were excluded, because these days were considered to be the period needed to become familiar with the vehicle. The first two days of the treatment phase were excluded, too, because they were considered to be needed to become familiarised with the distraction warning system. All inferential statistics were computed with an alpha level of 0.05.

Table 2 The number of days in the baseline phase and in the treatment condition per participant (first row), the chronological number of days per phase which were included in each phase (second row), and the total number of days per phase included in the analyses (third row). The first two days in each condition were discarded as adaptation period.

Participant Number	# days baseline (phase 1)	# days treatment (phase 2)	# days treatment (phase 3)	# days treatment (phase 4)	# days treatment total
10	11 (3–11) 9	 3–8 6	 9–14 6	 15–20 6	20 (3–20) 18
11	13 (3–13) 11	 3–8 6	 9–14 6	 15–21 7	21 (3–21) 19
12	13 (3–13) 11	 3–8 6	 9–14 6	 15–21 7	21 (3–21) 19
13	10 (3–10) 8	 3–8 6	 9–14 6	 15–20 6	20 (3–20) 18
14	12 (3–12) 10	 3–4; 7–11 7	 12–19 8	 20–26 7	24 (3–4;7–26) 22
15	10 (3–10) 8	 3–8 6	 9–14 6	 15–21 7	21 (3–21) 19
16	8 (3–8) 6	 3–6 4	 7–11 5	 12–15 4	15 (3–15) 15

3 Trip statistics

In this chapter an overview is given of the trip lengths and speed distributions for the different drivers and weeks. For the general trip statistics the first two days of the baseline and treatment phase were not filtered out.

Trip length is calculated based on the logged odometer data. Under normal operation, each time the ignition was switched off for longer than one minute a new log file was generated. Each separate log file is considered as one separate trip, no matter how much time passed between switching off the ignition and switching it on again. Therefore, the maximum trip length is in principle limited by the size of the tank of the car.

For each trip the first logged odometer value was subtracted from the last logged odometer value of the same trip in order to determine the logged distance of the current trip. The last odometer value of the previous trip was also subtracted from the first odometer value of the current trip, which resulted in the distance that was not logged in between those trips. This distance was added to the current trip for the total distance per trip, as it was assumed that in most cases the distance not logged was a result of the booting computer. A number of trips were compared to video recordings for validation of this hypothesis, and it proved to be true for a large majority of trips. However, especially at times when the computer did not boot correctly on the first attempt, a driver might already have finished a short trip before the computer had completed booting and initiating the log file, which therefore never was recorded. The not logged distances of those short trips are wrongly merged with the next recorded trips.

The figures in this chapter are presented on the same scale between participants, in order to make them easily comparable. In some cases this led to rather short bars, however.

3.1 Participant 10

Participant 10 drove altogether 3,033.5 km distributed across 310 trips, resulting in an average trip length of almost 10 km. In fact, however, she took many very short in-town trips of below 3 km, and several longer trips, most of them 20 km in length, between work and home. In total she drove on 31 days, which results in almost 100 km per day on average. How the number of trips and several other variables were distributed for this participant in both the *baseline* and the *treatment* phase on trips above and below a length of 3 km can be seen in Table 3.

In the same table it is also presented which percentage of the total distance was logged and which percentage was lost due to computer booting and rebooting. For slightly above 10 % of the total distance no data are available at all. For the short trips this percentage is much higher, because the booting phase of the computer lasted for about one minute when booting correctly, resulting in a higher percentage of the whole trip for the shorter trips than for the longer ones. Additionally, in some cases the computer did not boot correctly at the first trial, such that 5 min passed until the hard reset enforced a reboot. In this case, depending on speed, up to about 5 km passed before data were logged.

Table 3 Statistics about distances and number of trips for Participant 10.

baseline	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	913.2	133.8	1,047.0	12.8	113
above 3 km	861.2	109.4	970.6	11.3	72
below 3 km	52	24.4	76.4	31.9	41

treatment	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	1,806.1	180.4	1,986.5	9.1	197
above 3 km	1,705.3	136.9	1,842.2	7.4	107
below 3 km	100.8	43.5	144.3	30.1	90

Figure 1 shows which distance respectively time the participant spent in different speed intervals. In Sweden the speed limit in urban areas is 50 km/h, on small rural roads it is 70 km/h, on larger rural roads it is 90 km/h, and on the motorway it is 110 km/h. The participant's distribution across the different speed intervals in week four deviates from the other weeks to the effect that the participant drove faster than 110 km/h for a substantial percentage of time during week 4, whereas she did not drive that fast at all during the other weeks. Speeds above 110 km/h are indicative of motorway driving, which differs from rural road driving in many ways. A closer analysis of the data showed that the participant took a trip of around 500 km round trip on the motorway during that week, whereas she did not use the motorway at all during the other three weeks.

The distribution across the different speed intervals is relatively similar in weeks 1 through 3. The interval [90 km/h; 110 km/h] is also overrepresented in week 4.

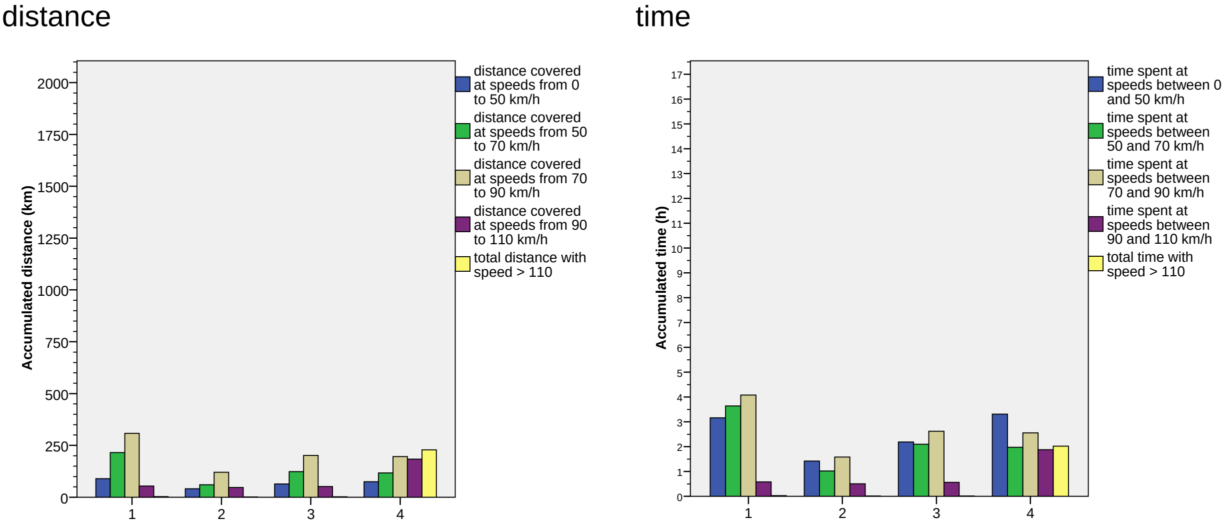


Figure 1 Distribution of speed per distance and per time for the baseline phase and three separate phases during treatment.

3.2 Participant 11

Participant 11 drove altogether 5,437.1 km distributed across 172 trips, resulting in an average trip length of 31.6 km. In total he drove on 34 days, which results in 160 km per day on average. During the *baseline* phase the average trip length was much longer than during the *treatment* phase. How the number of trips and several other variables were distributed for this participant in both the *baseline* and the *treatment* phase on trips above and below a length of 3 km can be seen in Table 4.

Table 4 Statistics about distances and number of trips for Participant 11.

baseline	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	3,095.6	42.8	3,138.4	1.3638	64
above 3 km	3,081.9	40.2	3,122.1	1.2876	51
below 3 km	13.7	2.6	16.3	15.9509	13

treatment	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	2,146.4	152.3	2,298.7	6.6255	108
above 3 km	2,116.1	140.6	2,256.7	6.2303	81
below 3 km	30.3	11.7	42	27.8571	27

This participant drove much more during week 1 than during the other three weeks, and he spent the major part of his driving during this week at speeds above 110 km/h. This participant generally engaged in a substantial amount of motorway driving. During week 2 he drove comparatively little, and the interval [90 km/h; 110 km/h] is under-represented during that week as compared to weeks 3 and 4. Those two weeks are relatively similar in the distribution of speed (Figure 2).

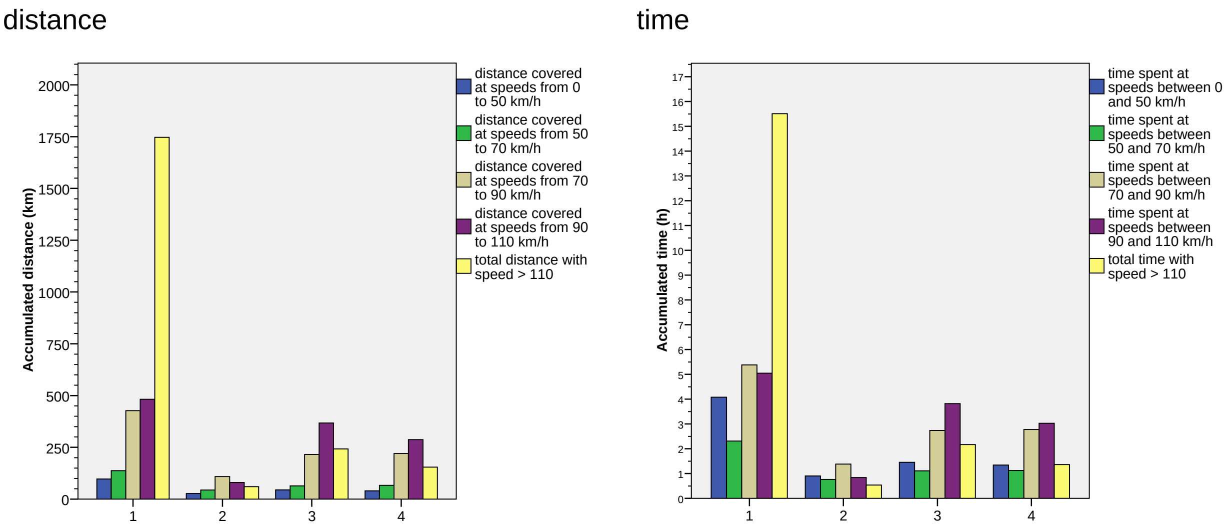


Figure 2 Distribution of speed per distance and per time for the baseline phase and three separate phases during treatment.

3.3 Participant 12

Participant 12 drove altogether 1,541.8 km distributed across 118 trips, resulting in an average trip length of 13.1 km. In total he had the car for 37 days, which results in 41.7 km per day on average. The average trip length was approximately equal for both phases. How the number of trips and several other variables were distributed for this participant in both the *baseline* and the *treatment* phase on trips above and below a length of 3 km can be seen in Table 5.

The participant shared the car with his girlfriend, because they car-pooled to his work, from where she then continued to her workplace. The girlfriend drove approximately 225 km during the *baseline* phase and approximately 300 km during the *treatment* phase. Her trips were excluded from the data analysis.

On several occasions the particpant wore a headband or a cap and had the collar of his jacket turned up, which impaired eye tracking. When he wore a headband with reflecting stripes, both eye and head tracking were reduced to zero.

Table 5 Statistics about distances and number of trips for Participant 12.

baseline ***(both drivers)***	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	847.9	85.9	933.8	9.199	75
above 3 km	825.2	79	904.2	8.737	59
below 3 km	22.7	6.9	29.6	23.3108	16

baseline ***(only intended driver)***	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	654.3	54.1	708.4	7.6370	50
above 3 km	640.2	49.4	689.6	7.1636	39
below 3 km	14.1	4.7	18.8	25	11

treatment ***(both drivers)***	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	993.4	146.6	1,140	12.8596	92
above 3 km	976.7	126.3	1,103	11.4506	73
below 3 km	16.7	20.3	37	54.8649	19

treatment ***(only intended driver)***	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	734.2	99.2	833.4	11.9030	68
above 3 km total	726	89.9	815.9	11.0185	58
below 3 km total	8.2	9.3	17.5	53.1429	10

Compared to the other participants this participant drove very little. He spent about half of his driving time at speeds below 50 km/h and drove a lot in urban areas. The distribution across the different speed intervals is fairly equal for all four weeks (Figure 3).

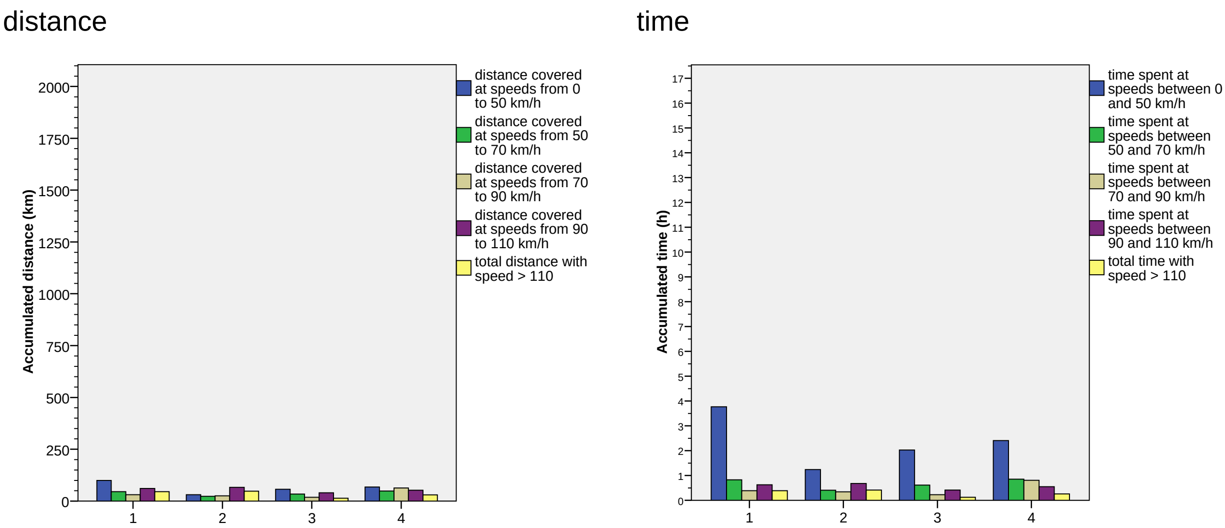


Figure 3 Distribution of speed per distance and per time for the baseline phase and three separate phases during treatment.

3.4 Participant 13

Participant 13 drove altogether 6,586.6 km distributed across 279 trips, resulting in an average trip length of 23.6 km. In total he drove on 30 days, which results in 220 km per day on average. During the *baseline* phase the average trip length was about 4 km longer than during the *treatment* phase. How the number of trips and several other variables were distributed for this participant in both the *baseline* and the *treatment* phase on trips above and below a length of 3 km can be seen in Table 6.

Table 6 Statistics about distances and number of trips for Participant 13.

baseline	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	2,060.4	76.5	2,136.9	3.58	80
above 3 km	2,039.5	69.7	2,109.2	3.3046	59
below 3 km	20.9	6.8	27.7	24.5487	21

treatment	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	4,042.4	407.3	4,449.7	9.1534	199
above 3 km	3,972.8	367.7	4,340.5	8.4714	130
below 3 km	69.6	39.6	109.2	36.2637	69

This participant drove a lot on motorways during all four weeks. The distribution across the different intervals is fairly equal across the four weeks except for week 1, where the interval [0 km/h; 50 km/h] is slightly overrepresented.

distance time

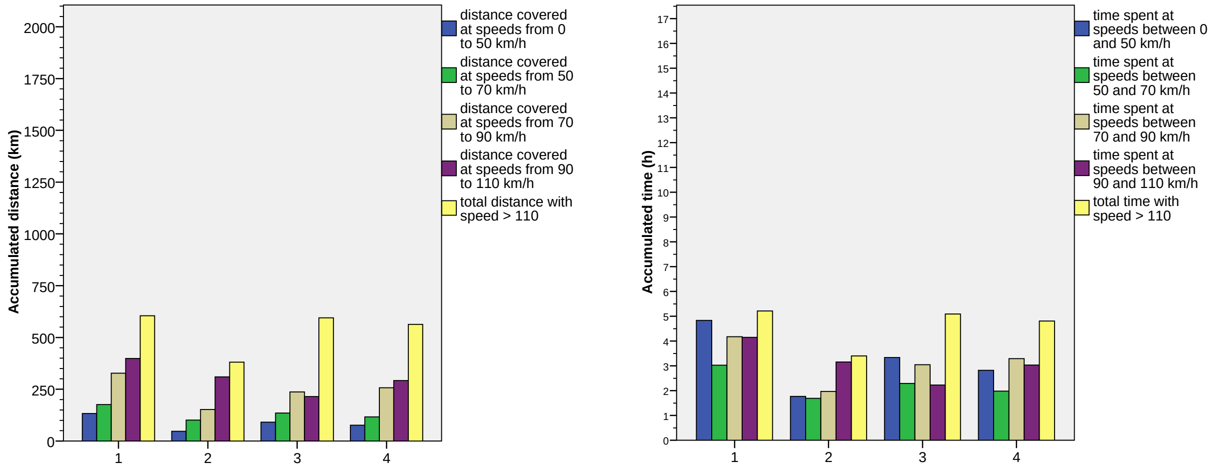


Figure 4 Distribution of speed per distance and per time for the baseline phase and three separate phases during treatment.

3.5 Participant 14

Participant 14 drove altogether 7,407.5 km distributed across 368 logged trips. During one longer trip in the fourth week the data acquisition system was unstable and restarted after about five minutes for the whole duration of the trip. This resulted in a number of short trips in the log, instead of one long trip, which inflated the trip number somewhat, also in Table 7.

During the treatment phase, the sponsor of the project scheduled a demo event for stakeholders and the press and required the car to be on site. Therefore the treatment phase for this participant was interrupted on day 4 in the evening, when the car was picked up from the participant. It was returned to him on day 6 in the evening. Due to this the first week of the treatment phase was extended with two days (see Table 2).

For this participant the eye tracking did not work as well as for the other participants. In many instances gaze tracking was lost and only head tracking was available. The video logs showed that the driver had a tendency of leaning his head back on the head rest, therefore he often “squinted” at the road with relatively closed eyes, which made it hard for the eye tracker to detect glance direction. Furthermore, the participant very often touched his face with his hand, thus, obstructing parts of the face from the view of the cameras.

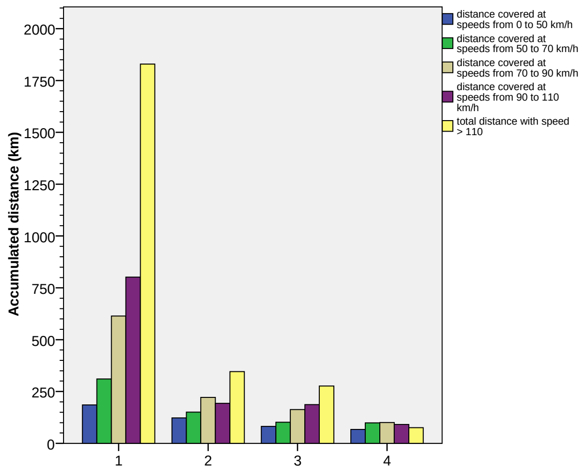
Table 7 Statistics about distances and number of trips for Participant 14.

baseline	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	4,172.8	117.3	4,290.1	2.7342	108
above 3 km	4,143	107.8	4,250.8	2.536	80
below 3 km	29.8	9.5	39.3	24.173	28

treatment	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	2,620.8	496.6	3,117.4	15.9299	260
above 3 km	2,540.7	449	2,989.7	15.0182	180
below 3 km	80.1	47.6	127.7	37.2749	80

This participant drove to Germany during the baseline phase, which is one reason for the much higher percentage of speeds above 110 km/h during the baseline phase than during the treatment phase. This trip explains, too, why the mileage during the baseline phase is that much higher than during the treatment phase (Figure 5).

distance



time

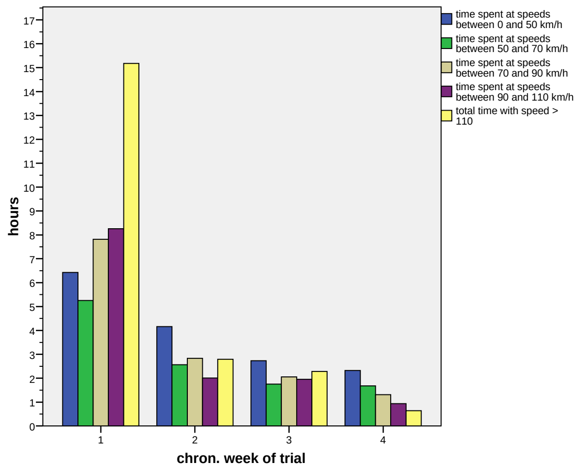


Figure 5 Distribution of speed per distance and per time for the baseline phase and three separate phases during treatment.

3.6 Participant 15

Participant 15 drove 2,568.4 km altogether with the experimental car. Due to the fact that the log system had become unstable, it was not possible without manual post-processing to determine how many trips were driven. The number of trips in Table 8 is therefore misleading, too. It becomes apparent that the data loss increased from the baseline phase to the treatment phase, which indicates that the computer became increasingly unstable over time. For the treatment phase one third of the distance was not logged, and in many instances the log was not only lost in the beginning of the trip, but also in the middle.

Table 8 Statistics about distances and number of trips for Participant 15.

baseline	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	740.3	149.2	889.5	16.7735	80
above 3 km	709.1	133.9	843	15.8837	56
below 3 km	31.2	15.3	46.5	32.9032	24

treatment	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	1,101.5	577.4	1,678.9	34.3916	251
above 3 km	999.9	496.7	1,496.6	33.1886	149
below 3 km	101.6	80.7	182.3	44.2677	102

The speed distributions are relatively similar for the different phases and weeks. In all cases the participant drove below 50 km/h for a substantial part of the time and distance driven. In week 4 she drove relatively little, but during this week most data were lost, too. Speeds above 110 km/h occurred almost only during the baseline phase (Figure 6).

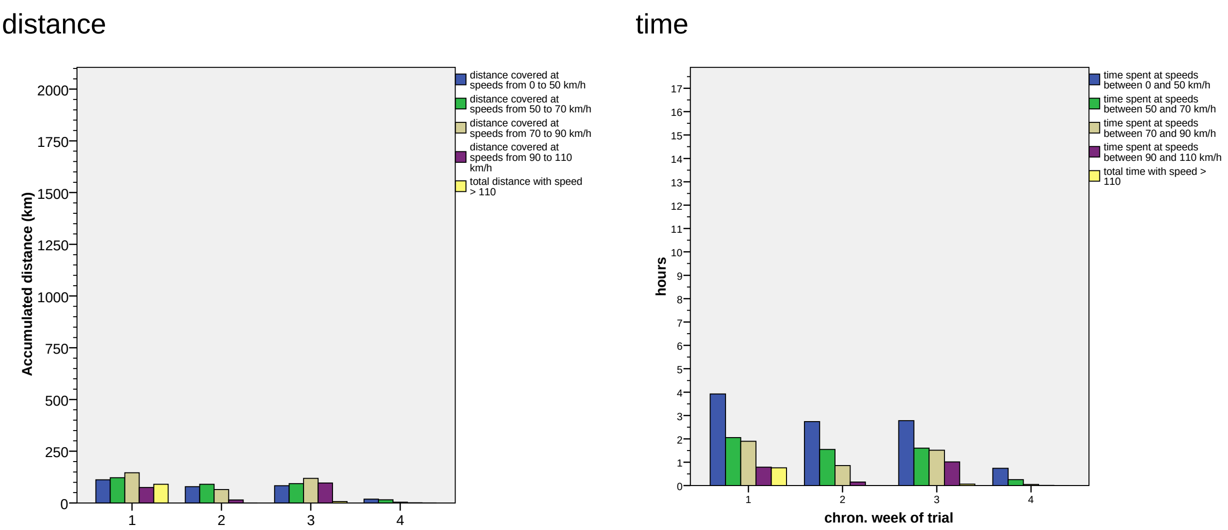


Figure 6 Distribution of speed per distance and per time for the baseline phase and three separate phases during treatment.

3.7 Participant 16

The car was returned to Saab for check-up due to the instability of the computer. Attempts were made to repair it, and it was decided to run Participant 16 as planned. Due to the approaching holidays and project end it was not possible to run a thorough trial to see whether the repairs had taken care of the computer crashes before the vehicle was handed over to the participant. It turned out that the problem had subsided somewhat during the baseline phase, but reappeared during the treatment phase.

Participant 16 drove 3,878.4 km in total, but for the same reasons as for Participant 15, namely the frequent computer crashes, it was not possible to determine how many trips she had made. During baseline driving data loss was limited to 6.6% of the total distance, but during the treatment phase about a third of the data were lost (Table 9).

Table 9 Statistics about distances and number of trips for Participant 16.

baseline	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	662.7	47.1	709.8	6.6357	47
above 3 km	652.1	45.2	697.3	6.4821	39
below 3 km	10.6	1.9	12.5	15.2	8

treatment	distance logged (in km)	distance not logged (in km)	total distance (in km)	percent not logged of total	number of trips
total	2,189.6	979	3,168.6	30.8969	143
above 3 km	2,150	957.1	3,107.1	30.8036	98
below 3 km	39.6	21.9	61.5	35.6098	45

The participant reported that she spent most of the time on the motorway, which is corroborated by the high percentage of speeds above 110 km/h. The speed distribution does not vary much across weeks, except that the percentage of speeds above 110 km/h is lower in the baseline phase than in the remaining weeks (Figure 7).

distance time

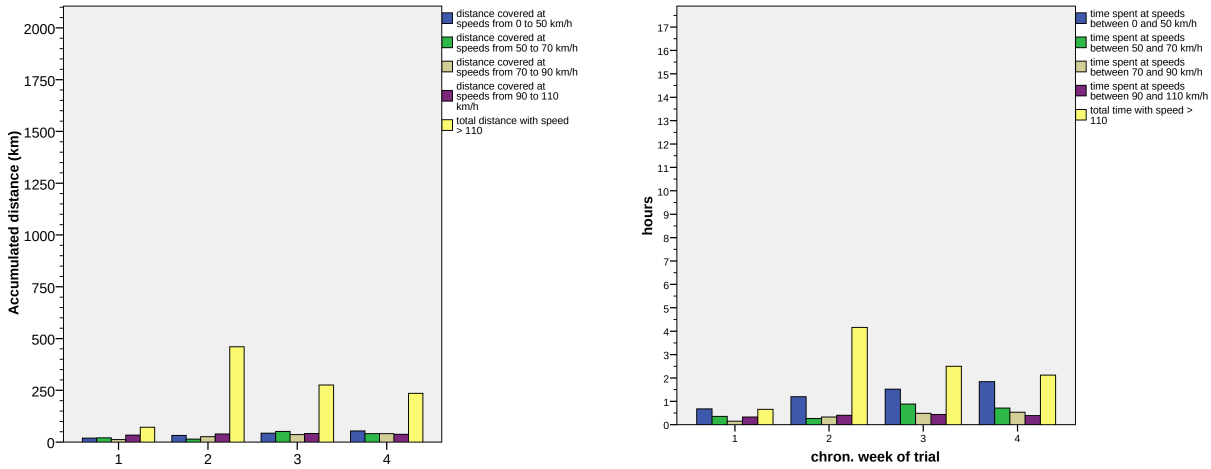


Figure 7 Distribution of speed per distance and per time for the baseline phase and three separate phases during treatment.

3.8 Comparison between participants

Mileage varied substantially between drivers, with Participant 12 having the smallest mileage and Participant 14 having driven almost five times as far within the same period of time (Table 10). Some participants drove the same route very often, because they mostly used the car for going to their workplace and back, while others drove many different routes. Three participants rarely drove above 110 km/h at all, whereas the other four spent a substantial part of their driving time at high speeds. Some participants showed relatively similar patterns across the four weeks of driving, while others had both very different mileages and different speed distribution patterns over the weeks.

Data loss was somewhat higher if the drivers took many small trips as compared to several longer ones. From the treatment phase of Participant 14 onwards it can be seen that data loss increased, then it fell again after the repairs before Participant 16 started, only to increase to a very high level again for the same participant's treatment phase.

Table 10 Overview over mileage and percent data loss for baseline and treatment phase for all participants.

	10	11	12	13	14	15	16
base (loss)	1,047 km (12.8 %)	3,138 km (1.4 %)	708 km (7.6 %)	2,137 km (3.6 %)	4,290 km (2.7 %)	890 km (16.8 %)	710 km (6.6 %)
treat (loss)	1,987 km (9.1 %)	2,299 km (6.6 %)	833 km (11.9 %)	4,450 km (9.2 %)	3,117 km (15.0 %)	1,679 km (34.4 %)	3,169 km (30.9 %)
total mileage	3,034 km	5,437 km	1,541 km	6,587 km	7,407 km	2,569 km	3,879 km

3.9 Discussion of trip statistics

The driving patterns of the participants varied both between and within participants, in the latter case between different weeks. Obviously, these different driving patterns reflect that drivers in the general population actually are very different from each other. This strong variation, together with the limited number of participants, complicates analysis and might cast doubts on some conclusions. For the most part the analyses will be made within subjects, but still considering the variation across subjects.

During field tests like this one, resembling FOTs with a relatively short driving time, special occurrences in driving patterns can have a relatively strong impact on the results. In this study one of the participants drove from Sweden to Germany and back in the baseline phase, which both led to a very high mileage during the baseline phase as compared to the treatment phase, and, more importantly, had some influence on the speed distribution during his driving. During the time when the study was conducted, the Swedish speed limits were 50 km/h, 70 km/h, 90 km/h and 110 km/h practically without exception. In Germany however, the speed limits are based on "even numbers", and in some cases on the motorway there is no speed limit at all. Generally the driving environment is quite different in those two countries.

Another participant drove to Stockholm and back to Linköping, a total distance of about 400 km, during the third treatment phase, and it was only on that occasion that she used the motorway. During the other three weeks she only drove on country roads and urban roads. Therefore the speed distribution as well as the driving environment was considerably different between this part of the treatment phase and the remaining time.

The other participants did not exhibit too varying patterns during the time they had the experimental vehicle, but this factor definitely has to be considered in future studies. Both a larger number of participants and longer driving times can help reducing impacts of that kind. Other possibilities are to control the driving environment better, or to classify road types and analyse data from different road types separately. This would, however, probably also require a larger number of participants to guarantee enough relevant events on each road type.

In the present study the participants were run serially, because only one passenger car was available. This was done in a country with marked seasonal variations, both when it comes to weather and hours of daylight. Driving in summer with very long daylight hours and in most cases good road friction differs a lot from driving during the winter months, when it is dark for the most part of the day, and when road conditions can be of varying friction. On some road types the winter maintenance includes ploughing and salting, whereas other roads can be covered with packed snow. Icy roads can occur on a regular basis during winter. Thus, the participants in the present study dealt with very different environmental conditions, which definitely varied between participants, but which also varied within participants. The winter during which the data were logged was relatively mild with only little snow, which was advantageous from the point of view of the study.

However, in general in Sweden it is not unlikely that during one week the weather is mostly sunny, while it rains a lot during the next week. This could influence the participant's behaviour more than the investigated safety system and should ideally be taken into account, either through logging and controlling for weather in the analyses, or by running enough participants in such a design that will most likely equal out possible confounding weather factors.

Generally, if unplanned influences like weather, type of road or hours of daylight cannot be randomised by a large enough number of participants and driving time, it would aid the interpretation of the results if matching situations could be found for analysis. One possibility would be to try to find a road segment that was frequented by all participants at least a certain number of times. If this proves impossible, a certain class of road might be used as selection criterion, possibly in combination with certain weather conditions. Any such special selection involves a substantial amount of more or less manual analysis, however, which did not fit into the budget frame of the present project. Therefore the only matching that was done for part of the analyses in this project was speed based. For some computations the results were split according to speed intervals that spanned 20 km/h. Often the speed interval of 0 to 50 km/h was excluded, because no warnings were given below 50 km/h, and all speeds above 110 km/h were subsumed into the same group. This was done to restrain the environmental variation within each group to some extent. Speeds of 110 km/h and above would occur mostly on motorways, speeds between 90 km/h and 110 km/h are very typical for relatively straight and well-built rural roads that can cover rather long distances. Speeds between 70 km/h and 90 km/h can often be found on smaller and more curvy rural roads, as well as on slightly more demanding sections on the bigger country roads, for example in junctions. Speeds between 50 km/h and 70 km/h are typical for suburban arterial roads and curvy and small rural roads. This classification takes into consideration that drivers often keep speeds that lie slightly above the posted speed limit. No cross-checking with the actual environment via GPS trace was conducted for the present analyses. The data are available, however, and could probably be matched to the Swedish national road database, which would not only provide information on the road type, but also on speed limits, number of lanes, road signs and a host of other variables. Obviously, in built-up areas the GPS log has to be quite accurate in order to achieve a correct match.

With respect to the analysis of the number of trips, it would have been recommendable to analyse the time and the distance that passed between the end of one log file and the beginning of the next. This would have aided in determining actual trip length. This was not done, because there is no readily available information on how long a break between two “driving sessions” may be in order to consider two data files to belong to the same trip. It might also be the case that additional information like a new goal should be considered when determining which unit is a trip. Thus, for simplicity reasons, and because this was not the focus of the study, it was decided to use a simple and easily applicable criterion for trip duration. Due to the computer crashes, however, not all files started with turning on the ignition and ended with turning it off.

4 Distraction warnings

The data were pre-processed in the following way before the effect of the warning system on warning frequency was analysed: All trips taken during the first two days with the car were excluded, because it was assumed that the participants needed that time to get used to the car. The trips during the first two days with the system activated were excluded, too, due to the participants' get used to the system. Additionally, all trips whose logs were shorter than 3 km were excluded, because during many of those short trips the participants rarely exceeded 50 km/h, the limit at which the distraction system began to work, and because those short logs were considered to be too idiosyncratic for a meaningful analysis. Finally, all trips during which the glance direction quality while driving did not exceed zero were excluded, because this indicated heavy problems with the eye tracking. This phenomenon occurred very rarely in general, and most often during very short trips.

In this chapter the frequency of distraction occurrences over time in general and in relation to different speed intervals is discussed.

Distraction occurrences are those events for which the algorithm determines that the driver's attention buffer is empty. Three different types of warnings were registered, which are *direct warnings*, *inhibited warnings* and *indirect warnings*.

Direct warnings are those that are issued to the driver immediately when the attention buffer becomes empty. The driver is considered to be distracted.

Inhibited warnings are occasions during which the attention buffer is empty, but no warning is issued either due to a speed below 50 km/h, activated direction indicators, the driver's braking or moving the steering wheel a lot, the last warning having been initiated less than 15 seconds ago or due to other inhibiting factors (Kircher, K., Kircher, A. & Claezon, F., 2009).

Indirect warnings occur when a filter for inhibition is active at the time when the attention buffer is emptied, but the filter for inhibition is deactivated while the attention buffer still is empty. Then the warning is issued immediately when the filter is deactivated.

During the baseline phase the same rules applied. In order to prevent the warnings from reaching the driver, the plug that connected the seat vibrator to power was pulled. This physical measure is not considered to be an inhibiting factor as defined above, because it does not have anything to do with the algorithm, but only with the experimental design.

The duration since the last warning was computed in a way that for each warning it was determined how many seconds had passed since the last warning had been given. Each warning was treated as one case, regardless of its chronological position within one trip. The first warning in each trip was excluded, because no meaningful time since the last warning could be computed. The warnings were always sorted into the speed interval at which the warning occurred.

For the analyses presented below only direct warnings were considered. The reason for this is that preliminary data analyses showed that the indirect warnings often should have been inhibited as well, especially when they occurred directly after the inhibition ceased due to vehicle data, like speed increasing to above 50 km/h or the driver's letting go of the brake. This is clarified with the help of an example:

When a driver leaves a roundabout, he often looks into the direction of the road he is going to drive onto, such that his gaze direction is not forward, but in the direction of the road he wants to continue on. In smaller roundabouts this direction may be located outside of the FRD. Often the speed in smaller roundabouts with stronger curves is below 50 km/h, though, therefore a warning for not looking at the FRD is inhibited. It happens, though, that the driver accelerates on his way out onto the other road, while still in the curve and still focusing his gaze into the direction of the remainder of the curve. If he exceeds 50 km/h during that phase, the speed inhibition is lifted. Then an indirect warning will be issued immediately if his attention buffer is still empty, even though the warning is not appropriate in this case. Analogous situations occur for braking. In the General Discussion under Section 11.6 ways to improve the AttenD algorithm are suggested, which also take this factor into consideration.

About one third of the indirect warnings occurred in the speed interval [50 km/h; 70 km/h], another third occurred at speeds above 110 km/h, and the remaining third occurred about equally distributed across the two intervals in the middle speed zones. Most of the direct warnings, which are computed by subtracting the number of indirect warnings from all warnings, are given at high speeds, whereas in the interval [50 km/h; 70 km/h] slightly fewer direct warnings are given than in the intervals in the middle speed zones (Table 11).

An analysis of the indirect warnings showed that in most of the cases they occurred about 15 s after the last issued warning, which indicates that they were given, because the inhibition that applied in the interval between [0 s; 15 s] after the start of the last warning was released, and the attention buffer was zero at the time. In the speed interval [50 km/h; 70 km/h] there was a tendency for longer intervals since the last warning was issued. This indicates that another main reason for indirect warnings in this speed interval was that the participants accelerated and exceeded 50 km/h while the attention buffer equalled zero. The number of indirect warnings was about 10% of all warnings for most of the participants. For Participant 12 the percentage of indirect warnings lay at 40%, and for Participant 14 the percentage lay at almost 30%, which is substantially higher than for the other participants. The actual number per participant and speed interval can be found in Table 11.

The average time that passed since a warning had been given is the content of this chapter. For the analyses the cases were split into the factors “week” and “speed category”. Week 1 represents the baseline, and week 2-4 represent the treatment condition. Exactly which days are included in which week can be found in Table 2. The warnings were sorted into different speed categories according to the speed at which the warning occurred. The speed at which the preceding warning was given was not considered.

The boxplots presented below are vertical boxplots. The boundaries of the box are first and third quartile. The median is identified by the middle line in the box. The length of the box is the interquartile range (IQR). Values more than three IQR’s from the end of a box are labeled as extreme (*). Values more than 1.5 IQR’s but less than 3 IQR’s from the end of the box are labeled as outliers (O).

Table 11 The number of indirect distraction warnings and the total number of distraction warnings per participant and speed interval, including the "silent warnings" in the baseline phase.

Interval	10		11		12		13	
	indirect	all	indirect	all	indirect	all	indirect	all
50–70	12	87	24	125	95	203	13	80
70–90	9	120	17	238	17	69	3	73
90–110	7	51	24	238	28	90	20	127
> 110	3	51	37	491	36	85	10	200
Total	31	309	102	1,092	176	447	46	480

Interval	14		15		16		Total (10-16)	
	indirect	all	indirect	all	indirect	all	indirect	all
50-70	174	477	7	68	4	17	329	1,057
70-90	136	557	3	41	4	13	189	1,111
90-110	110	451	6	38	2	21	197	1,016
> 110	233	883	1	19	9	79	329	1,808
Total	653	2,368	17	166	19	130	1,044	4,992

4.1 Participant 10

A bug in the distraction detection software was found first after Participant 10 had completed her run. The bug affected 31 of the trips for this participant. They were spread approximately equally over the four weeks during which she had the car. Most of those trips were shorter than 3 km. All affected files were excluded from the data analysis.

Table 12 Number of direct warnings given per week per speed category.

Interval	1	2	3	4	total
50–70	46	6	6	17	75
70–90	59	5	12	35	111
90–110	10	3	1	30	44
110+				48	48
Total	115	14	19	130	278

Figure 8 shows the time in minutes that passed between warnings at different speeds per week with the experimental car. Table 12 indicates how many of those warnings were issued for each of the conditions, thus providing an estimate for the reliability of the mean.

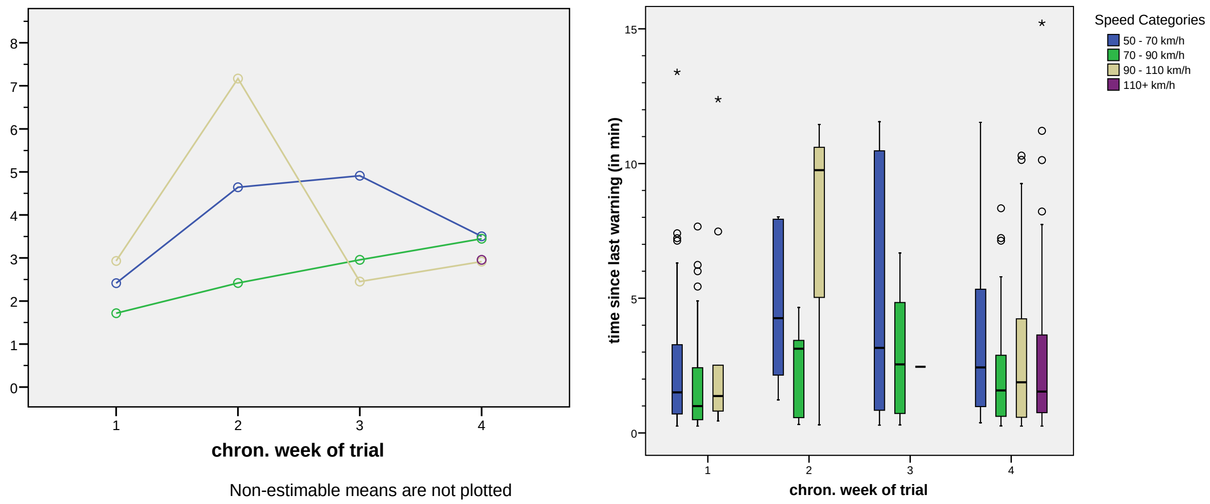


Figure 8 Estimated marginal means of the time in minutes since the last distraction warning was given per week per speed category (left) and boxplot of the time in minutes since the last distraction warning per week per speed category (right). Some extreme outliers are cut off for readability.

During the baseline week and during week 4 the participant drove substantially more and received many more warnings than during the other weeks. It also shows for those two weeks that the warning frequency does not vary much across speed categories. In weeks 2 and 3 where the mileage was lower and fewer warnings were received, the average frequencies vary more across speed categories, but it is likely that this is a result of the small number of observations.

The boxplot in Figure 8 shows, too, that the interquartile range is larger for the two middle weeks, during which only a small number of warnings was given, therefore mean values as those in Figure 8 should be viewed with caution. An analysis of variance with the factors week and speed category did not show any significant differences ($F(12, 265) = 1.18$).

4.2 Participant 11

For this participant all data files could be used, because the bug that had corrupted some of the files for Participant 10 had been fixed. The mileage of this participant was quite high, as was the number of warnings that he received.

Table 13 Number of direct warnings given per week per speed category.

Interval	1	2	3	4	total
50–70	45	19	12	25	101
70–90	90	31	40	60	221
90–110	81	11	70	52	214
110+	347	23	55	29	454
Total	563	84	177	166	990

This participant covered many miles, especially during the baseline phase. He received a substantial number of warnings in almost each of the phases and speed intervals, which makes the results relatively reliable.

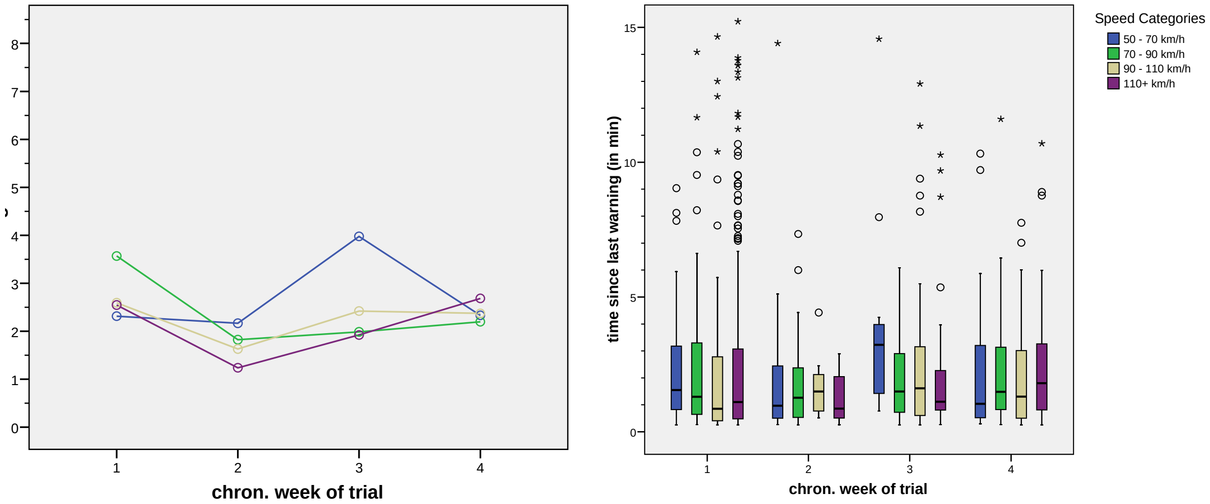


Figure 9 Estimated marginal means of the time in minutes since the last distraction warning was given per week per speed category (left) and boxplot of the time in minutes since the last distraction warning per week per speed category (right). Some extreme outliers are cut off for readability.

The mean intervals between two warnings are relatively equal across the different speed categories and across the four weeks (Figure 9). It appears that the mean duration between warnings decreased from the baseline phase to the first week in the treatment phase before returning to approximately the original level. No significant effects of the two factors week and speed interval were found, however ($F(15, 974) = 0.97$). The boxplot shows that the interquartile range for the intervals between warnings was approximately equal across speed categories and weeks, with the median having a tendency of being closer to the 25th quartile (Figure 9). For almost all boxes there are a substantial number of outliers and extremes, indicating that the variation in the time between warnings is quite big.

4.3 Participant 12

This participant had a relatively low mileage. He spent most of his driving time at lower speeds, where he also received most distraction warnings. In general, the number of distraction warnings per phase per speed category is relatively low for this participant, which indicates that mean values and distributions do not need to be very reliable (Table 14).

Table 14 Number of direct warnings given per week per speed category.

interval	1	2	3	4	total
50–70	37	23	25	23	108
70–90	13	9	12	18	52
90–110	20	13	18	11	62
110+	25	14	5	5	49
Total	95	59	60	57	271

The participant received a warning about once every 1.5 to 2 minutes on average (Figure 10). Neither speed nor phase influenced the duration between the warnings significantly (F(15, 255)=.93).

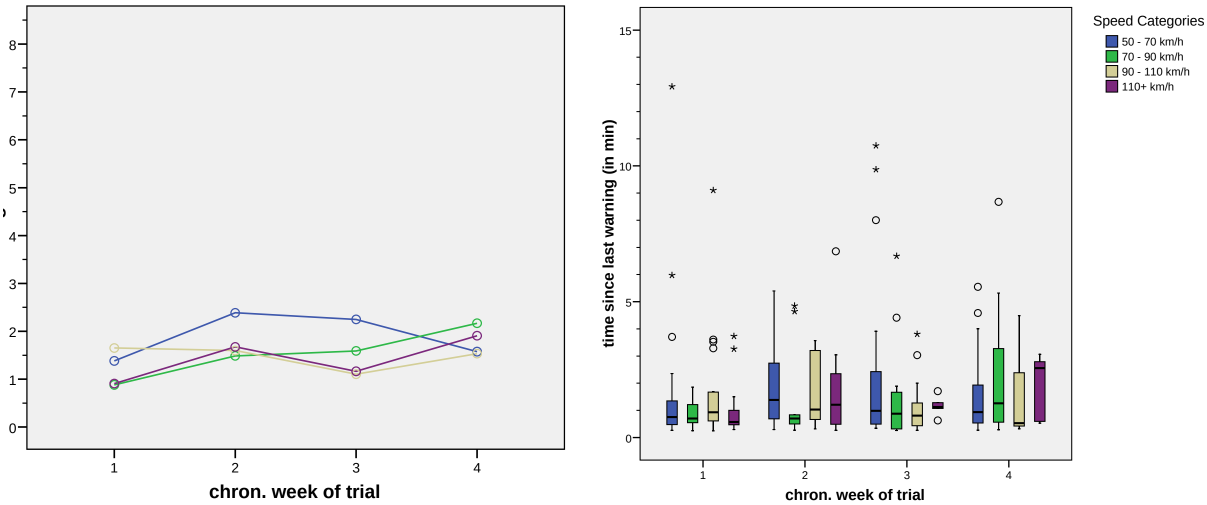


Figure 10 Estimated marginal means of the time in minutes since the last distraction warning was given per week per speed category (left) and boxplot of the time in minutes since the last distraction warning per week per speed category (right). Some extreme outliers are cut off for readability.

In the boxplot it can be seen that the median for the time between warnings lies at around 1 minute in most cases. Even though the two middle quartiles are relatively concentrated for this participant, the overall variation is quite substantial, and it has to be kept in mind that the number of observations within each quartile is rather small in many of the cases.

4.4 Participant 13

This participant had a high mileage, and a comparatively low number of warnings (Table 15). Due to the high mileage, however, it seems appropriate to assume that the average number of warnings per phase per speed category is reasonably representative for this participant.

Table 15 Number of direct warnings given per week per speed category.

interval	1	2	3	4	total
50–70	22	20	14	11	67
70–90	15	20	18	17	70
90–110	24	52	9	22	107
110+	62	32	48	48	190
Total	123	124	89	98	434

In Figure 11 the estimated mean durations between two warnings are displayed for Participant 13. In the baseline condition a warning would have been issued every third to fifth minute approximately, depending on the speed category. In the first treatment week the participant received the warnings more frequently than during the baseline phase, except for the highest speed category, for which the warning frequency decreased. For the three lower speed categories the warning frequency decreased continuously from the first treatment week to the third treatment week. For the highest speed category the trend was exactly opposite. An analysis of variance showed that there was a significant interaction effect between the two factors speed category and week ($F(9, 418) = 2.84$). There was a significant main effect for week, too ($F(3, 418) = 2.74$), which would probably grow stronger if the highest speed category were removed from the data.

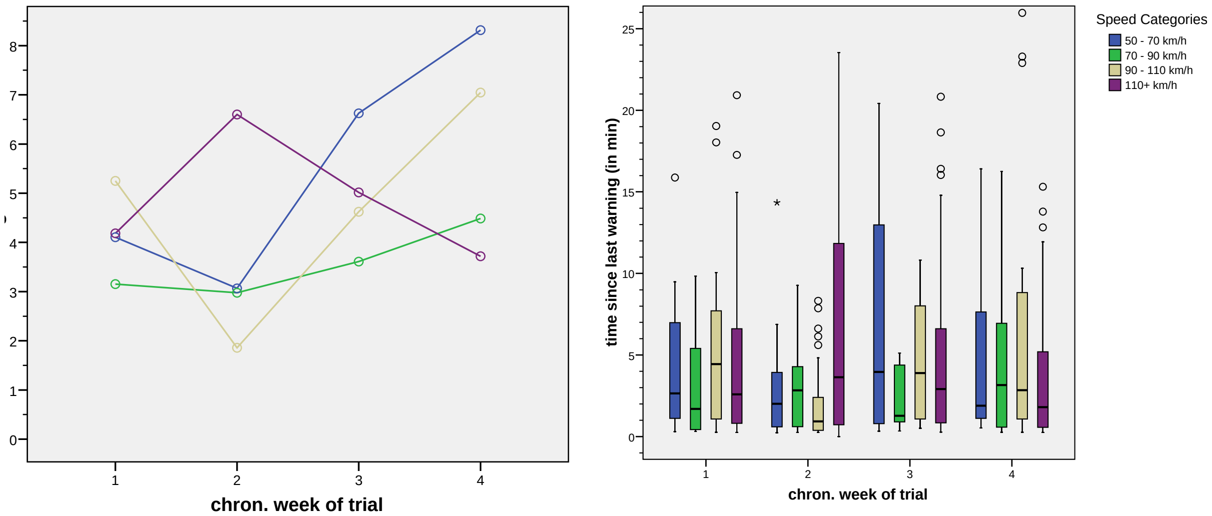


Figure 11 Estimated marginal means of the time in minutes since the last distraction warning was given per week per speed category (left) and boxplot of the time in minutes since the last distraction warning per week per speed category (right). Some extreme outliers are cut off for readability. Observe that the y-axis in the boxplot is not scaled exactly as for the other participants.

The boxplots show that for this participant the variance in duration between warnings was large (Figure 11). Generally the median lies closer to the 25^{th} than to the 50^{th} percentile.

4.5 Participant 14

This participant had a very high mileage, and he received at least 30 warnings per phase per speed category, which makes the results reasonably reliable (Table 16). Especially during the baseline phase the participant drove very much, which is also reflected in the number of warnings issued during this period.

Table 16 Number of direct warnings given per week per speed category.

interval	1	2	3	4	total
50–70	128	70	51	54	303
70–90	213	73	86	49	421
90–110	191	54	65	31	341
110+	454	74	79	43	650
Total	986	271	281	177	1,715

The mean time that passed between warnings lies between 1 and almost 3 minutes, depending on week and speed category. An analysis showed that both week and speed category had a significant effect on the duration between warnings; no interaction between the two factors was found (total: $F(15, 1699) = 3.43$; speed cat.: $F(3, 1699) = 5.21$; week: $F(3, 1699) = 9.45$). On average the time between warnings was about half a minute longer for the lowest speed category than for the other categories. Interestingly, the average warning frequency increased over time during the treatment weeks (Figure 12).

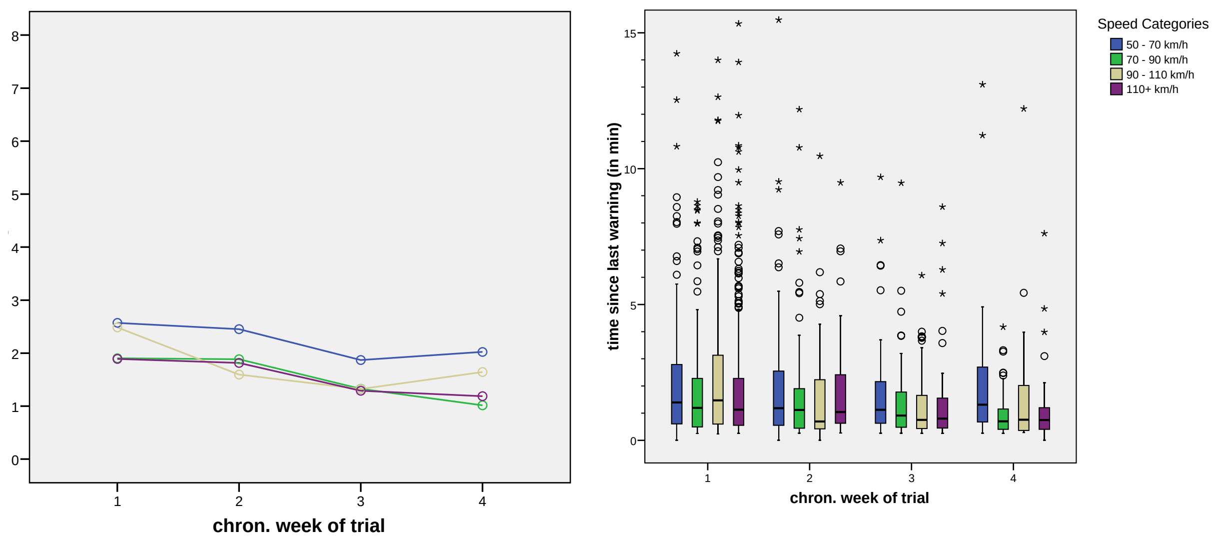


Figure 12 Estimated marginal means of the time in minutes since the last distraction warning was given per week per speed category (left) and boxplot of the time in minutes since the last distraction warning per week per speed category (right). Some extreme outliers are cut off for readability. Observe that the y-axis in the boxplot is not scaled exactly as for the other participants.

Also for this participant the box plot shows that the distribution of the time between warnings is quite large. Even though the centre quartiles lie within a range of two to three minutes, there are many outliers and extremes, some of them lie beyond 15 minutes.

4.6 Participant 15

As mentioned above, the data acquisition system became very unstable during the treatment phase of this participant, which led to very frequent computer crashes and restarts. This fact made it both impossible and meaningless to compute the time passed between two warnings. As described, the method to compute the average time between two warnings was to take every warning except for the first one as an event and to determine the time that had elapsed since the previous warning. This works well when each trip corresponds to one file. If a trip is split into many small files, however, it happens rarely that more than one warning can be found per file, therefore the computation cannot be made. Even if the files were concatenated the results would be misleading as the computer needs one to several minutes to boot, and during the booting time no warnings are issued.

Especially in week 2 and week 4 only very few warnings were issued at all, which does not allow any meaningful interpretation (Table 17).

Table 17 Number of direct warnings given per week per speed category.

interval	1	2	3	4	total
50–70	39	3	18	1	61
70–90	30	1	6	1	38
90–110	10	1	20	1	32
110+	18				18
Total	97	5	44	3	149

4.7 Participant 16

The computer problems described for Participant 15 were even worse for Participant 16, therefore no further computations except for listing the number of warnings that were issued per speed category and phase are made (Table 18).

Table 18 Number of direct warnings given per week per speed category.

interval	1	2	3	4	total
50–70	1	2	9	1	13
70–90		9			9
90–110	3	16			19
110+	6	62		2	70
Total	10	89	9	3	111

4.8 Discussion of distraction warnings

One commonality between the five participants for which data analysis was possible is the wide variation in inter-warning intervals. Practically all participants sometimes received warnings 15 s after the last warning, which is after the minimum possible time. For some participants it could take more than 20 minutes between two warnings, too.

For three out of five analysed participants no significant effect of week could be found. For Participant 13 the warnings became less frequent during the treatment phases than during the baseline phase. The interaction effect with speed is difficult to explain, however. When driving above 110 km/h, the warning frequency decreased markedly between baseline and the first treatment week, but it increased afterwards and returned to baseline levels in the last week, whereas for the other three speed categories the curve went the other way round, with an increased warning frequency in the first treatment week, which gradually decreased below baseline levels. Still, the boxplot shows that the range of the time that passed between warnings varies widely in all speed categories and weeks, which makes supposed differences between mean values less meaningful.

There is no general consistent effect of speed on the warning frequency in time. The drivers reach the distraction criterion as defined by the algorithm approximately as often when they drive on small country roads as when they drive on a motorway. This implies, on the other hand, that drivers cover greater distances on the motorway between distractions, because speeds are higher. Generally it is more meaningful to use time as denominator than to use distance, because a secondary task usually takes approximately the same amount of time, independent of speed.

The inhibition of 15 s after a warning was set, because it was assumed that a driver who continues with his secondary task in spite of a warning would get irritated at a system that then warns again. The idea was to warn once, making a potentially truly distracted driver aware of the fact that he neglected to monitor the traffic. If this was self-selected and conscious behaviour, however, and if the driver insists on continuing to look away from the FRD, another warning just after the first one will not help more than the first warning. It was chosen to set the inhibition time to 15 s in reference to the 15-s-rule, which was investigated by Tijerina, Johnston, Parmer, Winterbottom and Goodman (2000). This somewhat controversial rule states, however, that a task that takes 15 s to complete while the car is stopped would be safe to perform while driving, obviously with intermittent glances to the road. Therefore, the task will probably take longer than 15 s while driving. Therefore, it might be argued that a warning inhibition of 15 s is too short for drivers determined to complete a secondary task in spite of the warning.

In general, an average frequency of one distraction warning per a couple of minutes is considered to be too high. It is very likely that drivers will not accept this, and acceptance is important for a warning system to have the desired effect. Several suggestions of how the AttenD algorithm could be modified with the goal to reduce the warning frequency, especially by reducing the number of false alarms, are presented in the General discussion in Chapter 11.6.

5 Glance distribution

In this chapter it will be analysed in which way the drivers distribute their glances across different zones in the vehicle, and whether this distribution changes when the distraction warning system is activated. In the literature there exist suggestions for performance indicators for glance distribution which are independent of a world model of the vehicle, such as Percent Road Centre (Victor, 2005; Victor, Harbluk & Engström, 2005) and the Standard Deviation of Gaze (Recarte & Nunes, 2000; Zhang & Smith, 2004). These indicators can be computed for any gaze data set within a coordinate system. Other performance indicators presented here are based on a subdivision of the vehicle into different physical zones, the world model of the car. These indicators are not established in the literature, but they were used here, because they pertain directly to the AttenD algorithm used for distraction detection. They are the field relevant for driving (FRD), and the gaze distribution for gazes off the FRD, which is analysed in the end of this chapter.

5.1 Background

The eye movement data were collected with the SmartEye Pro system. The gaze direction data and the nose direction data are each described as a three-dimensional vector with an x-, y- and z-coordinate. The x-value describes the movement in horizontal direction with negative values indicating eye movements to the right and positive values indicating eye movements to the left, because the coordinate system is seen through the “eyes” of the cameras. The y-value describes the movement in vertical direction, with negative values indicating downward movements and positive values indicating upward movements. A gaze is represented by a unit vector describing the direction of the gaze and by a reference point describing the position of the participant’s eyes. For simplicity, the reference point was assumed to be located in the origin, resulting in gaze values lying on a sphere with radius 1. A sample of gaze points collected during a short sequence of a randomly selected trip is plotted three-dimensionally in Figure 13.

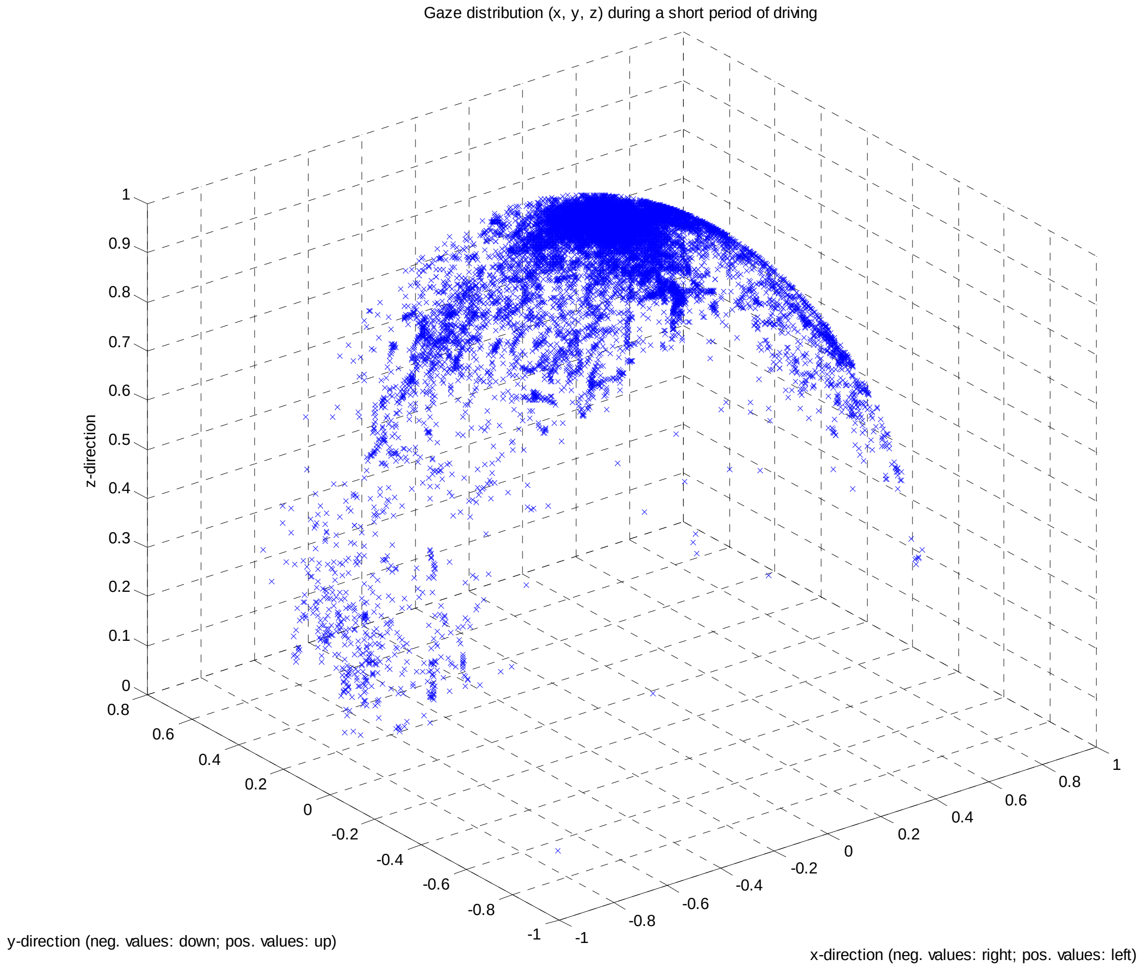


Figure 13 Gaze distribution of a short sequence of driving in a three-dimensional plot.

The spherical placement of the gaze points has some implications that have to be considered especially if only the x- and y-axes are used for computations. Some of the relationships will be explained with help of Figure 14. As mentioned above, the gaze points lie on a sphere with radius 1. The circle in the figure shows a cut through this sphere at y = 0, meaning that it is a top view of the coordinate system, with the driver's head in the centre.

By ignoring the z-values, it is implied that z = 0, and consequently that all values on the arc of the circle in Figure 14 are projected onto the x-axis (right part of the figure). This means that gaze points equally spaced along the arc of the circle will no longer be equally spaced when projected to the x-axis. The further to the side the driver looks, the more dense the distribution of x-values will be. The same effect is also demonstrated in the left upper quadrant of the circle in Figure 14, where it can be seen that equally spaced x-values correspond to very different sample areas along the arc of the circle.

For the analyses in this report equal spacing of the x-values was selected. The major motivation for this was simplicity. During driving, most glances lie within approximately [-0.5; 0.5]. For most cases, in which the head is rotated much more, no reliable gaze tracking is available, anyway. For the mentioned range the difference between the x-value from the 3D-system and the value where the gaze cuts the x-plane at z = 1 is not too big.

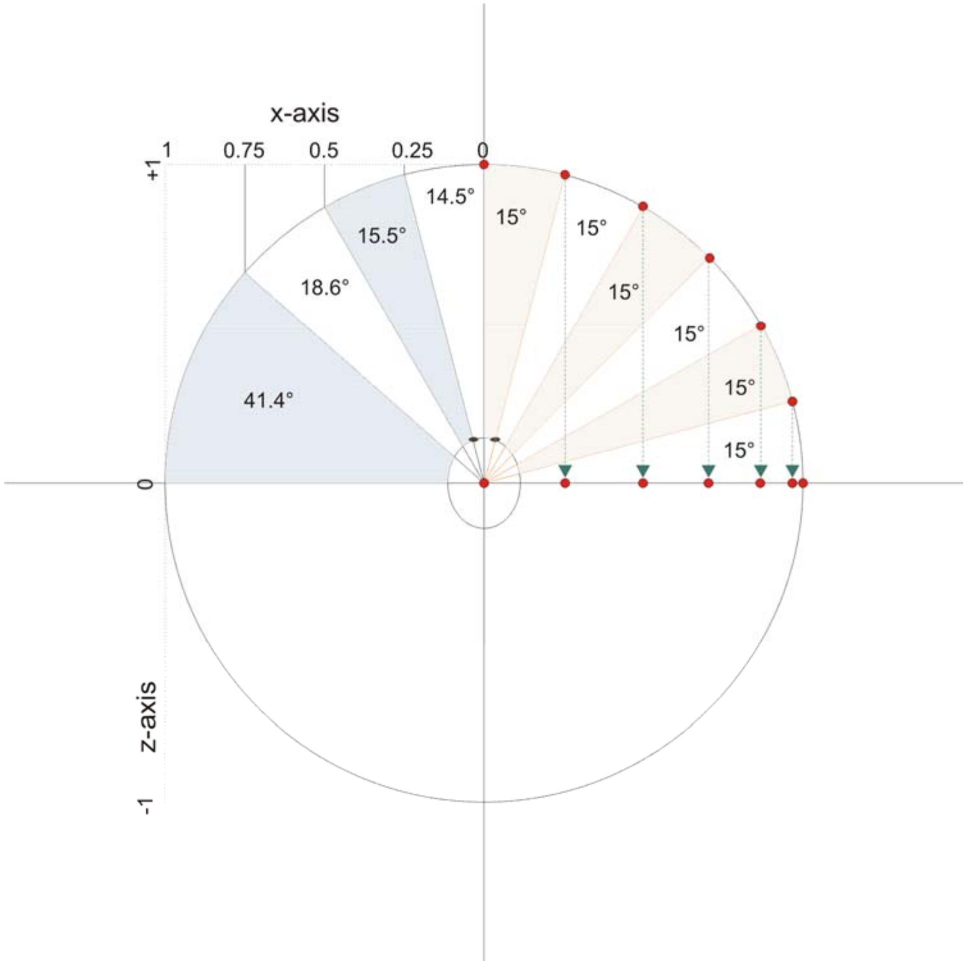


Figure 14 Top view of the driver's head, showing the x- and z-axes of the gaze coordinate system. On the left side of the frontal hemisphere the x-axis is split into equal portions, indicating the resulting gaze angles, and on the right side of the frontal hemisphere the x-values associated with equal gaze angles are shown.

5.2 Field Relevant for Driving (FRD)

One goal of the distraction warning system was to keep the driver's gaze within the field relevant for driving (FRD). The gaze is considered to be in this field when the driver looks through a window, not at a mirror, and the angle between straight ahead and the gaze direction is smaller than 45° (cf. Figure 15). It was hypothesised that the percentage of glances into the FRD is higher in the treatment condition than in the baseline condition, because the warning system should limit the drivers' glancing around.

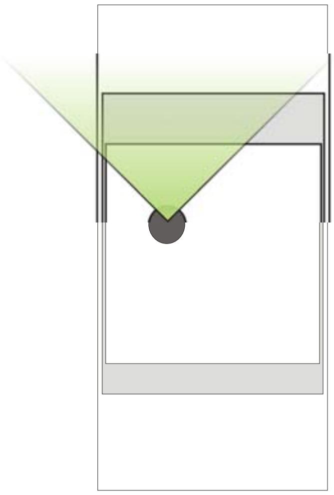


Figure 15 The green area depicts the Field Relevant for Driving (FRD), seen from above. Further requirements are that the gaze falls through one of the windows and not onto a mirror.

Together with the gaze direction itself a quality value for gaze direction was logged. This value reflects the strength of the edge between iris and sclera and is normalised to the range [0.0, 1.0]. The normalised value 0.0 corresponds to the 1st percentile of all collected quality values of the current trip, and 1.0 corresponds to the 99th percentile of all those values. This implies that the algorithm needs a few seconds before the quality value stabilises (for reference and further discussions see page 64 of SmartEye AB, 2007). This was disregarded for the computations described below, because a few seconds were considered to be negligible when compared to the average trip length.

Head tracking quality, also called nose direction quality, is either 0, 0.5 or 1, depending on whether no camera, only one of the two cameras or both cameras have head tracking.

Trips taken during the first two days of the baseline condition and during the first two days of the treatment condition were excluded, because this period was considered to be needed for adaptation to the vehicle respectively to the functionality of the distraction warning system. For the analyses presented here all speeds were included.

The glance distribution within and outside of the FRD for the baseline was then compared to the glance distribution in the treatment condition. Several different selection criteria for the glances included in the comparison were applied. Table 19 shows the percentage of the remaining gaze points after eliminating all gaze points in which the reported gaze direction quality (see above) was lower than 0.25. This was done to exclude gaze points with unreliable or no tracking. In Table 20 the values are given for all cases in which the quality criterion of 0.25 was met, or else, head tracking was present with a quality value of at least 0.5, and the speed exceeded zero. By this situations in which the driver could look around freely, because the vehicle was not moving, were removed from the data set before the percentages were calculated. Additionally, cases with head tracking only were included. For all drivers except for Participant 12 a substantial increase in the percentage of valid tracking cases can be observed after this procedure.

Table 19 Percentage of gaze cases for which the gaze direction quality exceeded 0.25 and the nose direction quality was equal or larger than 0.5. Treatment phase data are reported per week and in total.

	P10	P11	P12	P13	P14	P15	P16
baseline	74.5 %	79.8 %	69.0 %	81.9 %	74.9 %	81.9 %	82.0 %
treat 1	73.3 %	70.6 %	72.1 %	80.1 %	72.3 %	83.0 %	83.5 %
treat 2	73.7 %	78.8 %	76.5 %	81.2 %	70.1 %	81.8 %	81.1 %
treat 3	77.3 %	75.7 %	77.2 %	82.9 %	67.5 %	82.7 %	79.0 %
total treatment	75.2 %	76.1 %	75.6 %	81.4 %	70.4 %	82.4 %	81.1 %

In Table 19 the percentage for good enough eye tracking is given. Whenever eye tracking is present at least one camera has head tracking, too. In Table 20, however, the percentage for either good enough eye tracking (>0.25) or good enough head tracking (≥0.5) is provided, indicating when any kind of tracking was present as long as the vehicle was in motion.

Table 20 Percentage of gaze cases for which the gaze direction quality exceeded 0.25 or the nose direction quality equalled or exceeded 0.5 while the speed was strictly greater than 0 km/h. Treatment phase data are reported per week and in total.

	P10	P11	P12	P13	P14	P15	P16
baseline	98.7 %	96.2 %	75.6 %	97.1 %	90.4 %	96.8 %	97.4 %
treat 1	98.9 %	96.2 %	83.2 %	97.0 %	89.6 %	97.8 %	97.3 %
treat 2	98.7 %	96.1 %	87.4 %	96.7 %	88.2 %	95.9 %	96.6 %
treat 3	98.8 %	96.0 %	85.6 %	96.8 %	85.6 %	94.6 %	96.3 %
total treatment	98.8 %	96.1 %	85.4 %	96.9 %	88.0 %	96.4 %	96.8 %

The rows which fulfilled the different quality requirements for gaze direction were then further processed with respect to whether the gaze for each instance fell into the FRD or not. Only the percentage of gaze instances that fell into the FRD for speeds excluding 0 km/h is reported in Table 21 for the baseline phase and for the treatment phase, because there is practically no difference between the percentage of gazes within the FRD when speeds at 0 km/h are included. Additionally, the percentage for each week of the treatment phase is reported. It has to be noted that all computations were made on raw gaze cases, without pre-processing the data to determine fixations and saccades.

Table 21 Percentage of gaze cases within the field relevant for driving (FRD) based on gaze cases for which the gaze direction quality exceeded 0.25 or the nose direction quality equalled or exceeded 0.5 while the speed was strictly greater than 0 km/h. Treatment condition data are reported per week and in total.

	P10	P11	P12	P13	P14	P15	P16
baseline	78.8 %	81.0 %	80.6 %	89.0 %	81.3 %	80.8 %	84.4 %
treat 1	80.9 %	81.5 %	81.9 %	87.0 %	81.4 %	81.0 %	81.1 %
treat 2	80.1 %	82.6 %	83.2 %	88.3 %	78.3 %	80.5 %	83.1 %
treat 3	80.2 %	80.7 %	85.0 %	89.4 %	76.2 %	81.5 %	80.8 %
total treatment	80.3 %	81.7 %	83.6 %	88.2 %	78.8 %	80.8 %	81.7 %

The data were filtered once more to base the statistics on only those cases in which a distraction warning would have been possible. This excludes all cases in which speed was below 50 km/h, as well as when the brake or the indicator were active, when the steering wheel was moved markedly, and when not more than 15 s since the last warning had passed. The percentage of gaze instances which fell onto the FRD for those data is shown in Table 22. The percentages are similar or a few percent points higher than in Table 21. This allows the conclusion that during the times when an inhibition is active the percentage of glances off the FRD is slightly higher than during times in which no inhibition is active. It does not say, however, whether the glances away from the FRD are of a different quality for those two conditions, meaning that the glances away during an inhibition are not due to distraction, whereas those that occur while no inhibition is active are.

Table 22 Percentage of gazes within the field relevant for driving (FRD) based on gaze cases for which the gaze direction quality exceeded 0.25 or the nose direction quality equalled or exceeded 0.5, while no warning inhibitions are active. Treatment condition data are reported per week and in total.

	P10	P11	P12	P13	P14	P15	P16
baseline	79.8%	81.5%	82.3%	90.1%	81.0%	81.8%	85.7%
treat 1	84.6%	83.4%	86.5%	88.5%	82.1%	83.5%	82.2%
treat 2	83.9%	83.7%	86.0%	90.6%	78.2%	82.1%	84.4%
treat 3	81.6%	82.0%	89.1%	91.3%	76.7%	84.3%	81.8%
total treatment	82.9%	83.0%	87.6%	90.2%	79.2%	82.8%	82.8%

In general it is not the case that the percentage of gaze cases within the FRD is higher during the treatment condition than during the baseline condition.

5.3 Percent Road Centre

Percent Road Centre (PRC) is a performance indicator that has been used by Victor (2005) to determine the level of attention. It was reported to be sensitive to visual task difficulty and also, albeit to a lesser extent, to auditory task difficulty (Victor, Harbluk & Engström, 2005).

Here it is hypothesised that the PRC is larger in the treatment phase than in the baseline phase, because the participants' eyes are directed forward due to the warning.

Victor et al. computed PRC by binning the gaze data which were determined to be fixations into 128 by 128 bins for a 120 by 120 degree portion of the data in the forward view in order to determine the mode, or most frequent gaze angle. The road centre was then defined as a circular area around that point with a radius of 8 degrees. It was noted, however, that using the complete gaze data, and not only fixations, would probably have yielded similar results (p. 181).

In the present study all gaze data which exceeded a gaze direction quality threshold value of 0.25 were binned into 100 by 100 bins for a 180 by 180 degree field of view for each separate trip of each driver (see Figure 16). The lower number of bins was used to make up for possibly decreased accuracy of the eye tracker in the used field setting. The binning occurred over the x- and y-axes only, which implies that bins lying further to the edges collect gaze points from a greater viewing angle than bins lying closer to the centre. As for the FRD computations the raw gaze cases were used, without computing fixations first.

A circular area with a radius of 8 degrees around the modal bin was then defined as the road centre. The PRC for each subset of trips for each participant was determined by weighting the trips with the number of valid gaze cases. Thus, a short trip would get a lower weight than a long trip in the final PRC, because each valid gaze case got the same weighting.

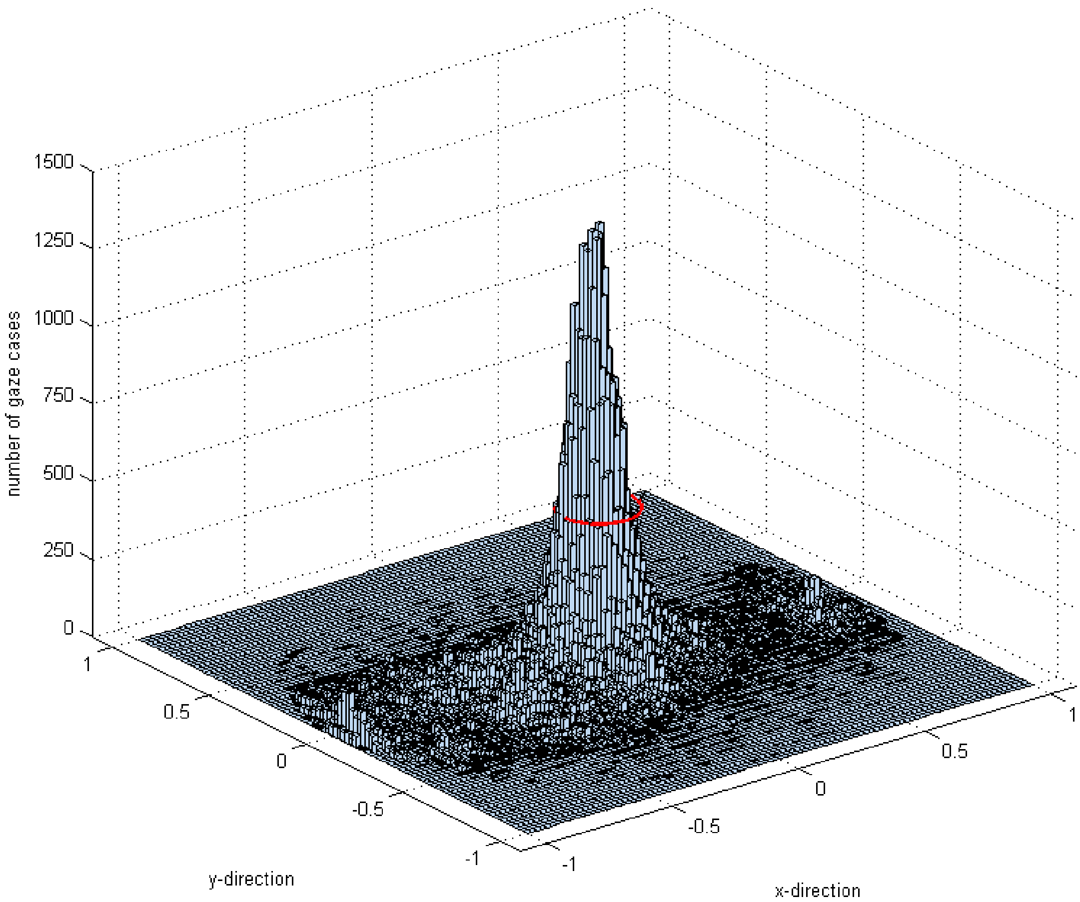


Figure 16 The distribution of gaze cases with a gaze direction quality >0.25 for one trip of one participant. The red circle indicates the “road centre”, according to PRC with a radius of 8°. X-direction is left-right, y-direction is up-down.

The PRC values for the seven participants are much more consistent within each participant than between participants (see Table 23). Participant 15 glances at the road centre for about 30% of the valid gaze cases, whereas Participant 13 reaches values of around 70%. Within participants, however, the difference in PRC over the weeks lies usually at around four or five percent. Only for participant 14 a larger change in percentage could be observed, with the PRC decreasing markedly from the baseline phase to the treatment phase. More detailed analyses showed that especially for Participant 14 the circle defining the road centre intersected with portions of the rear view mirror. This leads to the situation that mirror glances, which are not included in the FRD, are included into the PRC value.

Table 23 Percentage of Percent Road Centre (PRC) based on gaze cases for which the gaze direction quality exceeded 0.25 or the nose direction quality equalled or exceeded 0.5, while the speed was strictly greater than 0 km/h. Treatment phase data are reported per week and in total.

	P10	P11	P12	P13	P14	P15	P16
Baseline	50.3 %	52.2 %	36.7 %	71.6 %	49.8 %	33.6 %	53.8 %
treat 1	47.2 %	52.7 %	39.6 %	68.9 %	39.5 %	30.4 %	44.7 %
treat 2	48.6 %	54.2 %	39.4 %	70.1 %	38.2 %	35.8 %	53.6 %
treat 3	56.2 %	56.0 %	38.9 %	69.7 %	34.8 %	30.8 %	50.0 %
total treatment	51.6 %	54.6 %	39.3 %	69.6 %	37.7 %	33.2 %	49.1 %

An additional exclusion of all those cases during which a distraction warning inhibition was active yields a different picture (Table 24). The PRC values are in general several percent points higher when the data are filtered that way. This implies that the glances deviate away from the road centre much more when the warnings are inhibited. This goes in line with the findings for the FRD, but the result is more pronounced for PRC. No clear pattern can be seen for a change between baseline and treatment phases – for some participants the PRC value increases slightly for the treatment phase, while it decreases for others. Thus, the hypothesis propagating an overall PRC change due to the distraction warning cannot be supported.

Table 24 Percentage and absolute number of percent road centre (PRC) based on gaze cases for which the gaze direction quality exceeded 0.25 or the nose direction quality exceeded 0.5 while there are no inhibited warnings. Treatment phase data are reported per week and in total.

	P10	P11	P12	P13	P14	P15	P16
Baseline	62.2 %	55.0 %	44.6 %	78.9 %	54.4 %	40.7 %	61.1 %
treat 1	61.2 %	59.5 %	48.6 %	75.3 %	44.3 %	36.0 %	47.7 %
treat 2	59.4 %	58.0 %	48.6 %	77.6 %	44.1 %	41.5 %	59.8 %
treat 3	63.9 %	61.1 %	43.1 %	75.6 %	39.9 %	38.8 %	55.7 %
total treatment	61.9 %	59.5 %	46.1 %	76.2 %	43.1 %	39.5 %	53.6 %

The results from both the FRD percentages and the PRC values together indicate that approximately one third to half of the glances within the FRD fall outside the road centre. Only for Participant 13 the picture looks different. For him about 80% of the glances within the FRD fall inside of the road centre. The remaining glances, that are not inside the PRC, but still inside the FRD, are distributed across the window areas outside of the road centre circle.

5.4 Standard deviation of gaze

The standard deviation of gaze direction and of nose direction are computed on the respective x- and y-values measured (cf. Figure 14). Within each participant the standard deviations are very similar across the weeks (see Table 25 and Table 26). There is some variation between the participants, and again, Participant 13 is an exception with a noticeably low standard deviation of gaze. In general the standard deviation in x-direction, which is horizontal, is about twice as big as the standard deviation in y-direction, which is vertical. Except for the standard deviation in x-direction for Participant 14, who showed an increased standard deviation during treatment, no substantial increase or decrease could be found between the baseline condition and the treatment condition for this indicator.

Table 25 Standard deviation of gaze in the horizontal direction, based on all cases fulfilling the gaze quality criterion of >.25. Treatment phase data are reported per week and in total.

	P10	P11	P12	P13	P14	P15	P16
baseline	0.28	0.25	0.24	0.19	0.25	0.34	0.29
treat 1	0.28	0.25	0.24	0.20	0.28	0.34	0.28
treat 2	0.27	0.24	0.26	0.20	0.29	0.33	0.29
treat 3	0.26	0.24	0.23	0.20	0.30	0.31	0.29
total treatment	0.27	0.24	0.24	0.20	0.29	0.33	0.29

Table 26 Standard deviation of gaze in the vertical direction, based on all cases fulfilling the gaze quality criterion of >.25. Treatment phase data are reported per week and in total.

	P10	P11	P12	P13	P14	P15	P16
baseline	0.12	0.13	0.16	0.10	0.11	0.13	0.11
treat 1	0.12	0.13	0.15	0.11	0.12	0.14	0.11
treat 2	0.11	0.13	0.15	0.10	0.12	0.14	0.11
treat 3	0.10	0.14	0.14	0.10	0.12	0.15	0.12
total treatment	0.11	0.13	0.15	0.10	0.12	0.14	0.11

The standard deviation of gaze corresponds well to the PRC data. Those participants that keep their gaze more often in the road centre also show a smaller standard deviation of gaze, both in the lateral and the longitudinal direction. Just as for PRC, the standard deviation of gaze is smaller for head movements than for gaze movements.

5.5 Gaze distribution off road

The vehicle was subdivided into 15 different glance targets, so-called “zones” as described in Kircher et al. (2009). The zones are listed in Table 27. These zones were used for the eye tracker and the distraction detection algorithm AttenD. The eye tracker works 3-dimensional, meaning that a glance can cut through more than one zone. One example would be a glance at the left rear view mirror. The glance vector also cuts through the left side window. In those cases in which more than one zone is cut by the gaze vector, the transparency of the first zone determines which zone is coded as gaze target. If the first zone is transparent, only the second zone is coded. In the case of the mirror, for example, it is assumed that the driver wants to check the mirror and does not look at the window pane located in front of the mirror. If the first zone cut by the vector is not transparent, the first zone is coded. This would be the case for the centre rear view mirror, which is not transparent, therefore the windscreen behind the mirror is not a visual target.

For the analyses in this subchapter only those glances were considered that occurred when no distraction warning inhibition criteria were on.

Table 27 The 15 zones into which the vehicle is subdivided. The transparent zones are in italics.

#	name	comment
2	*windscreen*	transparent, part of windscreen in FRD, glances at the centre rear view mirror are not coded as windscreen
3	*right front window*	transparent, all glances that fall through the right window and on the right mirror are coded as right mirror only
4	*the left front window*	transparent, all glances that fall through the left window and on the left mirror are coded as left mirror only
5	right rear view mirror	glances that fall through the window onto the mirror are coded as mirror only
6	left rear view mirror	glances that fall through the window onto the mirror are coded as mirror only
7	centre rear view mirror	in front of windscreen, all glances that fall onto the mirror are coded as mirror only, not as windscreen
8	dashboard	logically behind speedometer, glove compartment and middle console, all glances that fall on speedometer, glove compartment and middle console are not coded as dashboard
9	speedometer	in front of dashboard, all glances that fall onto the speedometer are coded as speedometer only, not dashboard
10	middle console	in front of dashboard, all glances that fall onto the middle console are coded as middle console only, not dashboard
11	the glove box	in front of dashboard, all glances that fall onto the glove box are coded as glove box only, not dashboard
12	left front door	window glances do not belong to the door area
13	right front door	window glances do not belong to the door area
14	floor of the car	down on the floor
15	foot area	further up front, from approximately the knees downward to where the pedals are
16	roof	

In Figure 18 the distribution of glances off the FRD is presented per participant per phase. Each glance had equal weight, regardless of its length. The computation was made in the following way:

Whenever the AttenD algorithm registered that the driver directed his gaze out of the FRD the buffer decreased, as described in the chapter Introduction and background. When the driver looked back at the FRD the buffer increased again after a latency period of 0.1 s. Glance duration was computed as the time that elapsed from when the buffer started to decrease until the time when the buffer started to increase again, including the 0.1 s latency phase. This way the time during which the buffer remained at zero was included, like in the last and next to last glance in Figure 17.

All glances that were shorter than 0.2 s including the latency period of 0.1 s, and all glances with less than 80% of eye tracking above a quality value of 0.25 were excluded from the analysis.

In order to determine at which object the driver looked during a given glance, the zone with the highest percentage of gaze registrations during a given glance was determined. That is, a glance with 60% of the registrations in the rear-view mirror and 40% of the registrations in the windscreen was classified as a rear-view mirror glance. For zone determination only glances with an eye tracking quality of at least 0.25 were used.

For glances to the mirrors and to the speedometer the countdown of the buffer was delayed with one second in order not to “punish” drivers for traffic relevant glances off the FRD. Therefore, only glances to the speedometer and the mirrors that exceed 1 s are included in the zonal distribution of glances presented in Figure 18, as only those are considered not to belong to traffic relevant glances. This means that drivers who do not have any registered mirror glances can very well look at the mirrors, but they do so for shorter than 1 s.

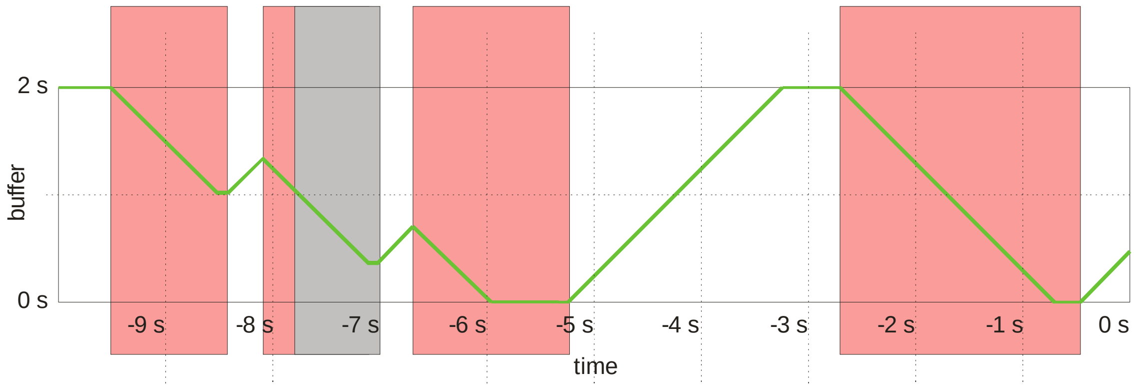


Figure 17 An example of a possible timeline of the attention buffer (green). Glances off FRD are marked in pink. The grey area indicates head tracking only.

In Table 28 it can be seen for which percentage of all glances off the FRD it was not possible to determine into which zone the driver looked, because the tracking quality did not meet the inclusion criterion of at least 80% eye tracking during that glance, or because the glance was shorter than 0.2 s. For Participants 12 and 15 more than half of the glances had to be discarded. For Participant 16 only the first treatment phase appears to be problematic. Otherwise data loss for glances off the FRD does not appear to deviate too much from the overall data loss due to low tracking (compare Table 19).

Table 28 Percentage of all glances outside of the FRD for which a zone classification was not possible due to insufficient tracking quality.

	baseline	treatment 1	treatment 2	treatment 3
Participant 10	22.0	22.5	21.9	16.2
Participant 11	30.0	26.3	27.5	20.7
Participant 12	73.6	74.4	75.4	73.7
Participant 13	29.1	27.2	31.4	33.8
Participant 14	27.4	35.7	34.8	35.0
Participant 15	51.3	53.7	50.3	45.8
Participant 16	31.1	48.9	26.8	33.6

In Figure 18 it can be seen that across participants and phases the part of the windscreen (number 2) that lay outside of the FRD received most glances of all glances outside of the FRD. The dashboard (number 8) is another area that was looked at frequently, except for Participant 12. For this participant a high percentage of glances to the roof (16) was logged. The speedometer (9), the middle console (10) and the left window (4) are looked at relatively frequently, but only very few gazes to the right window (3), the glove box (11), the floor (14 and 15) and the car doors (12 and 13) were registered. Of the three mirrors, the centre mirror (7) received the highest glance count, but again, only mirror glances lasting more than 1 s were included in the figure.

Generally the glance distribution for off-FRD glances looks very similar across the phases within each participant. The interindividual differences are not striking, either, except for Participant 12, whose distribution seems to deviate somewhat from the others'.

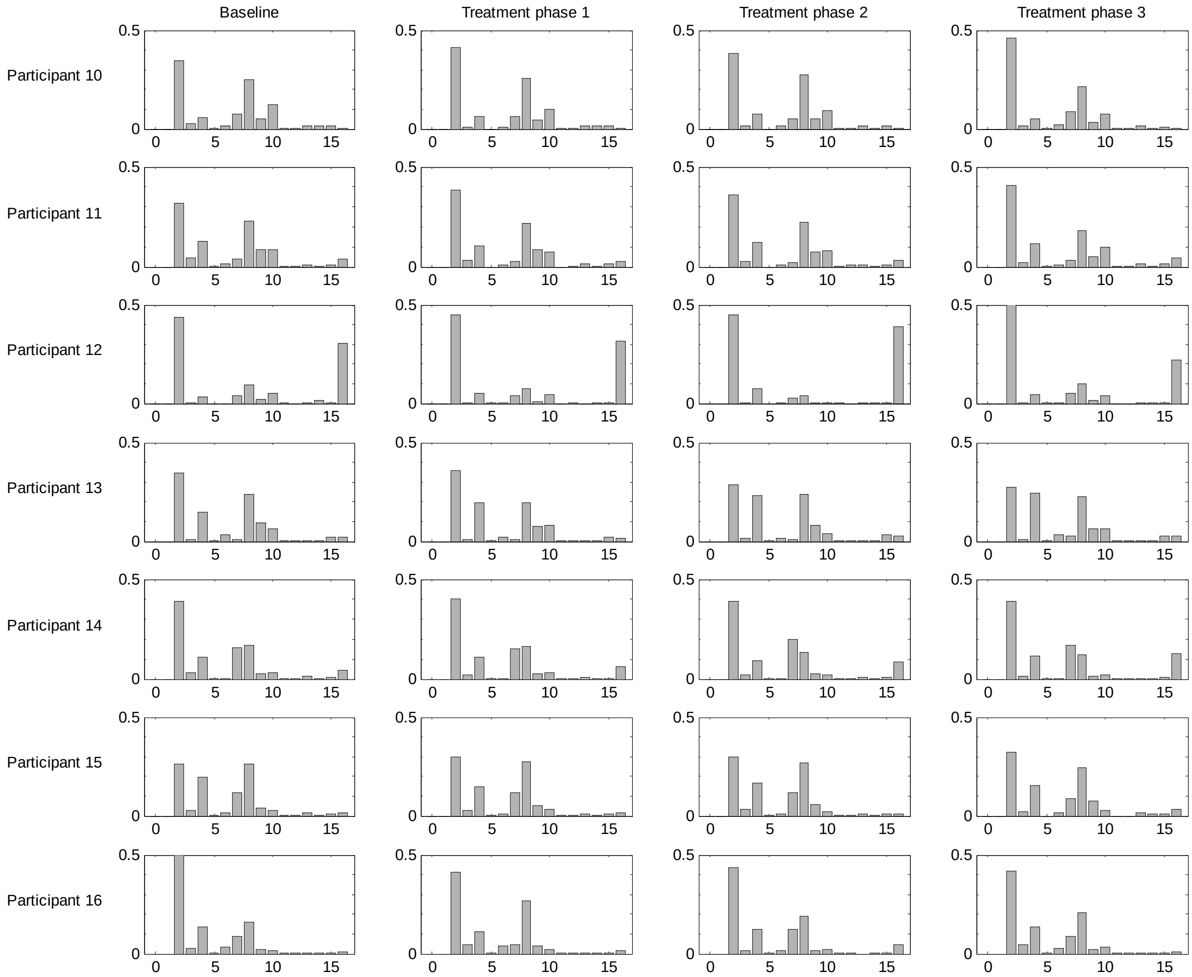


Figure 18 Percental distribution of off-FRD glances into the 15 zones of the car per participant per phase in per cent of total number of glances off-FRD (Zones are counted from 2 to 16).

In Figure 19 the zonal distribution is presented in relation to the total number of off-FRD glances. A full circle represents all glances, both inside and outside FRD, that occurred while no distraction warning inhibition was on. The size of the segment shows the percentage of glances outside the FRD, whereas the size of the “missing” segment shows the percentage of glances inside the FRD. The glances that were excluded due to failed tracking are not represented visually. The FRD values were computed based on gaze cases, as in the tables above. They could not be computed glance based, as the basis for the glance computations in this report is the AttenD algorithm, which only indicates one very long “glance” for those instances when the driver looks at the FRD. A fixation-based analysis would have allowed glance-based behaviour for glances inside the FRD, too.

The gaze cases registered as falling into the FRD are based on cases which had either eye or nose tracking. Nose tracking is enough to determine whether a glance is inside the FRD or not. The zone distribution, however, is only based on eye tracking, as this information is necessary in order to determine at which zone the gaze was directed. The representation assumes that there is no systematic relationship between the gaze falling onto one zone and its likelihood to be detected by eye tracking as opposed to nose tracking. It is likely, however, that such a relationship exists, which probably leads to the fact that zones with a smaller visual angle to road centre are overrepresented as compared to those with a larger visual angle to road centre.

Furthermore, in the figures presented here zone distribution is based on glance counts instead of on gaze case counts, because this was found to lead to a better indication of the actual gaze behaviour.

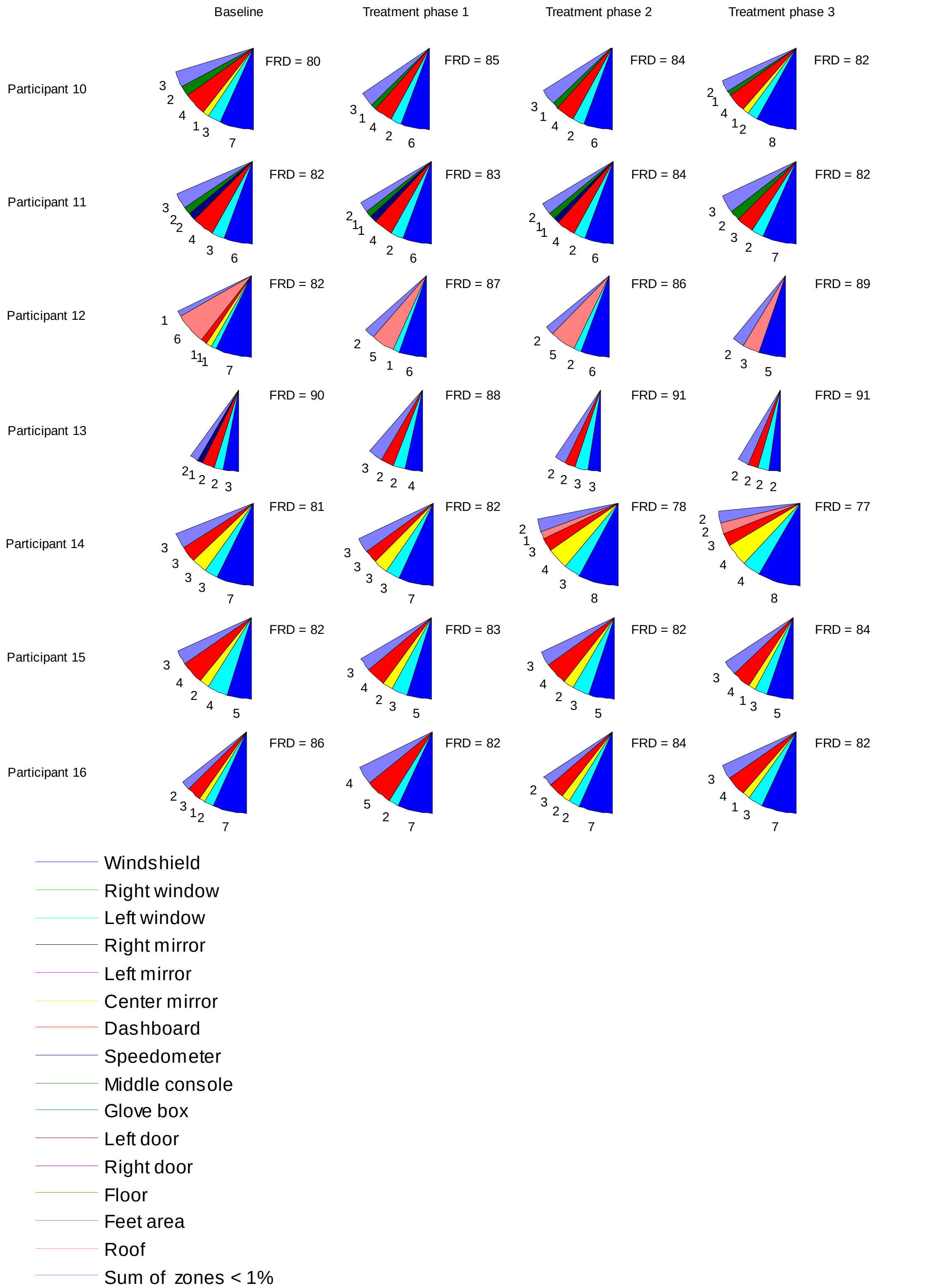


Figure 19 Percental distribution of off-FRD glances with at least 80% tracking into the different zones with the size of the sector representing the percentage of all gaze registrations outside of FRD, given that the complete gaze data both in- and outside FRD are represented by a full circle. The numbers represent the percentage of the total.

The glance distribution inside and outside the FRD, as well as into the different zones is much more alike within one participant for the different phases than within the same phase for different participants. Figure 19 contains the combined combination of Figure 18 and Table 21.

5.6 Discussion of glance distribution

The results in Table 21 and Table 22 show that drivers look away from the FRD for up to around 20% of their driving time. This includes, however, glances to windshield areas that lie outside of the FRD, prolonged glances to the mirror and prolonged glances to the speedometer. On the other hand, these data are based on only those gaze cases that have at least satisfactory nose tracking, while all the cases for which no tracking exists are excluded. As long as the vehicle was in motion, the average complete loss of tracking lay between 1.2% and 24.4% per phase per driver, but in most cases at about 4%. For Participant 12 the loss in tracking was much higher than for the other participants, followed by Participant 14. Especially for Participant 12 the percentage of head/nose tracking only was much smaller than for the other participants.

An analysis of the video tapes showed that Participant 12 often wore caps or head bands, sometimes with reflecting stripes on them that completely prevented tracking. He wore jackets with high collars, too, which covered the lower part of his face up to his nose, making tracking difficult and at times unreliable. Participant 14, on the other hand, had a habit of touching his nose with his hand and thereby obstructing parts of his face from camera view. This led to a higher percentage of tracking loss, but still kept the percentage of head/nose tracking only at a level comparable to that of the other participants.

Even though it is unlikely that all instances of lost tracking are due to the driver's turning away his head enough to render tracking impossible, there is a good chance that the drivers do not look at the FRD for a substantial percentage of the no-tracking cases. This reasoning is not valid for Participant 12 and 14, for whom the increased occurrence of loss of tracking could be explained by other phenomena like clothing and hand movements.

There is no significant change in the percentage of lost tracking between the baseline phase and the treatment phases. A decrease in lost tracking could have meant that drivers look more at the forward roadway, which would lead to smaller data loss tending to occur at extreme head angles. This would have needed to be taken into account when looking at the other analyses. As it is, it appears like there is no reason to assume a systematic shift in loss of tracking with and without a distraction warning system active.

The FRD was designed to be a somewhat simplified, but generally valid representation of where a driver was supposed to look during standard driving tasks. For an assessment of the AttenD algorithm used in this study, it would be necessary to evaluate the delimitations of the FRD thoroughly. For the study described here the FRD was set to be fixed, even though there is a built-in possibility to let it move laterally in conjunction with steering-wheel movements. This was meant to represent the curves of the road, which should determine the area relevant to the driver. It was supposed that while driving in a curve to the right the glance direction should be biased to the right, and the other way around. This feature had, due to the relatively large area covered by the FRD, only marginal influence, therefore it was not used for the present study.

As mentioned, the data show that a driver looked away from the FRD for 20% of the total driving time on average. In the 100-car study it was found out that looking away from the forward roadway for more than two seconds accumulated within a six seconds interval means an increased risk for crashes (Klauer, Dingus, Neale, Sudweeks & Ramsey, 2006). On average, during a six seconds interval, the drivers in the present study looked away for 1.2 seconds accumulated, which equals 20%. In the 100-car study, where eye gaze data were hand coded, the average time during which the gaze was directed off the forward roadway lay at 0.9 s accumulated for baseline driving. It is not clear whether the difference of 0.3 s on average between the two studies is due to the differing measurement techniques, the definition of "forward roadway" vs. FRD, or due to other reasons like the different traffic environment in the two studies. The value of 0.3 s is not huge, however, and the mean time off the FRD for any given 6-s-interval in the present study lies well below the value found for crashes in the 100-car-study, which lay slightly above 1.8 s.

The 20% outside of the FRD are a very rough measure, calculated across the complete driving time. This does not mean that for each 6 s, the driver looks away for 1.2 s. Rather, there will be 6 s intervals during which the driver does not look away from the FRD at all, and there will be some during which the driver reaches the 2 s criterion found in the 100-car study, which corresponds to 33.3% of the 6 s. Generally it can be said that a driver with a high total percentage of glances off FRD is more likely to arrive at high off-FRD values for selected 6 s intervals than a driver who has a low total percentage of glances off the FRD.

As mentioned above, for most participants about one third of the glances within the FRD do not fall into the road centre. Glances outside of the road centre but inside the FRD should indicate traffic relevant scanning behaviour, if the AttenD algorithm works correctly. Obviously, the algorithm does not have access to information about the surrounding traffic, therefore it is impossible to determine whether a fixation in a particular area at a certain time was traffic relevant or not.

Most drivers direct their glances at the road centre for about 35 to 50% of the time, but the individual variance between drivers is rather large. Here the PRC value was computed as one accumulated value of several days. No significant changes that resulted from the activation of the distraction warning could be found in the present data, when comparing entire phases. This means that possible effects of the system were not strong enough to show in data accumulated in such a rough way, but this does not mean that there is no effect of the system at all.

In other studies the PRC value was determined as running value per minute (e.g. Victor, Harbluk & Engström, 2005). This allows comparisons for situations in which a driver was distracted with baseline driving data. In the present study such a fine-grained comparison was not done due to budget limitations. A suggestion would be to compute the PRC for the last minute before a warning, for the first minute after a warning and for comparable baseline driving, that is, for the same trip, at comparable speeds, when no warning was issued. Also shorter durations are thinkable. An analysis of how the value changes from before to after the warning, and in which way this relates to baseline would allow conclusions about the effectiveness of the warning. The time until baseline behaviour is reached again could also be determined. For the present data this type of analysis is, however, complicated by the fact that so many warnings were issued. For this reason it might not be easy to determine what data belong to baseline and what data still belong to the time after a warning or before the next warning. Therefore it is recommended to perform this type of analysis after a change in the algorithm that decreases the number of issued warnings substantially.

It was discovered that the circle around the road centre in some cases intersected with the rear view mirror. It is a matter of discussion whether glances into the rear view mirror really should be considered “road centre”, regardless of how relevant they might be for driving. This issue will be dealt with more thoroughly in future analyses and publications.

Standard deviation of gaze could be a problematic indicator as it increases with external visual tasks relevant to driving, indicating active scanning of the environment, but also with visual distractions outside of the field relevant for driving. Analogous to the findings for percent road centre (PRC) it might be argued that an extraordinary decrease in standard deviation of gaze could be indicative of cognitive distraction. However, the distraction warning system reacts to visual distraction only.

The overall standard deviation of gaze was not influenced by the warnings (see Table 25 and Table 26), which is not particularly surprising. Only from the standard deviation of gaze it is not possible to determine whether the driver looks to the side due to a distraction or due to active scanning behaviour. Just as for the PRC values, a more fine-grained analysis would be necessary to be able to find more subtle effects of the distraction warning system.

The glance distribution for those off-FRD glances for which a zone could be determined shows more interindividual differences than intraindividual differences between the phases (Figure 18 and Figure 19). One commonality between the participants is that a substantial portion of the off-FRD glances was directed at those parts of the windscreen that were not part of the FRD. This might indicate that larger portions of the windscreen are, in fact, relevant for driving. This could be an explanation for the difference in the gaze percentage off the forward roadway found for the present study and in the 100-car study. It could, however, also indicate that drivers spend time looking at distracting objects outside of the vehicle. This cannot be determined without information about which objects were present in the traffic environment at the time of the off-FRD windscreen glances.

The dashboard is another area that drew around 20% of the off-FRD glances (Figure 18). Even though the speedometer had its own zone, the imprecision in tracking quality could lead to a number of speedometer glances being registered as dashboard glances. In this case they did not qualify for the one second latency period. There are other instruments on the dashboard, too, however, which might very well have received a number of glances. This figure is therefore not very surprising. There are a number of glances through the left window, which are probably part of either normal scanning behaviour or distractions outside of the car on that side of the road. Furthermore, just as for the dashboard and the speedometer, some left mirror glances might have been recorded as left window glances.

Only two participants had a noticeable number of glances at the middle console, and two other participants tended to direct longer glances at the centre mirror, such that it showed up in the off-FRD data. This underlines, again, that drivers have different glance patterns, and that it is very important for further analyses and further studies to consider the interindividual differences. For Participant 12 many glances to the roof were registered. This is most likely a result of the less accurate tracking for this participant, which was commented on before. He tended to wear garments that either decreased the accuracy of eye tracking or rendered tracking completely impossible.

To sum up, the usage of glance distribution based performance indicators on a rough level as was done here does not appear promising, and due to the reasons explained above, not very meaningful. Suggestions for further analyses will be given in the general discussion in Chapter 11.

6 Glance duration

In this chapter single glance duration, regardless of whether it resulted in an inattention warning or not, was analysed. Special emphasis was put on long glances. It was hypothesised that the warning system would reduce the occurrence of very long glances above 2 s. Those glances are generally considered to be most detrimental for traffic safety.

Due to the fact that glance duration is not normally distributed, that the data are quite inhomogeneous, and that the number of participants was low, no inferential statistics were computed. The analyses are therefore only on a descriptive level. They were made per participant and phase, as in the other chapters. Preliminary analyses showed that no substantial differences could be found between different speed categories, which is why all speed groups were considered together.

6.1 Extracting glance duration statistics

Information about glance occurrences was extracted from the time trace of the attention buffer, where start and stop times of each glance was calculated based on zero crossings in the buffers negative derivative. Since the size of the attention buffer is 2 s, a glance calculated in this manner can maximally be 2 s long. To circumvent this shortcoming, the duration of the glance leading up to the warning was stretched until the attention buffer started to increase again. If two glances were closer than 0.1 s in time (the latency period of the AttenD algorithm) they were merged into a single glance. All glances with less than 80 per cent tracking or with a duration of less than 0.2 s were omitted from further studies.

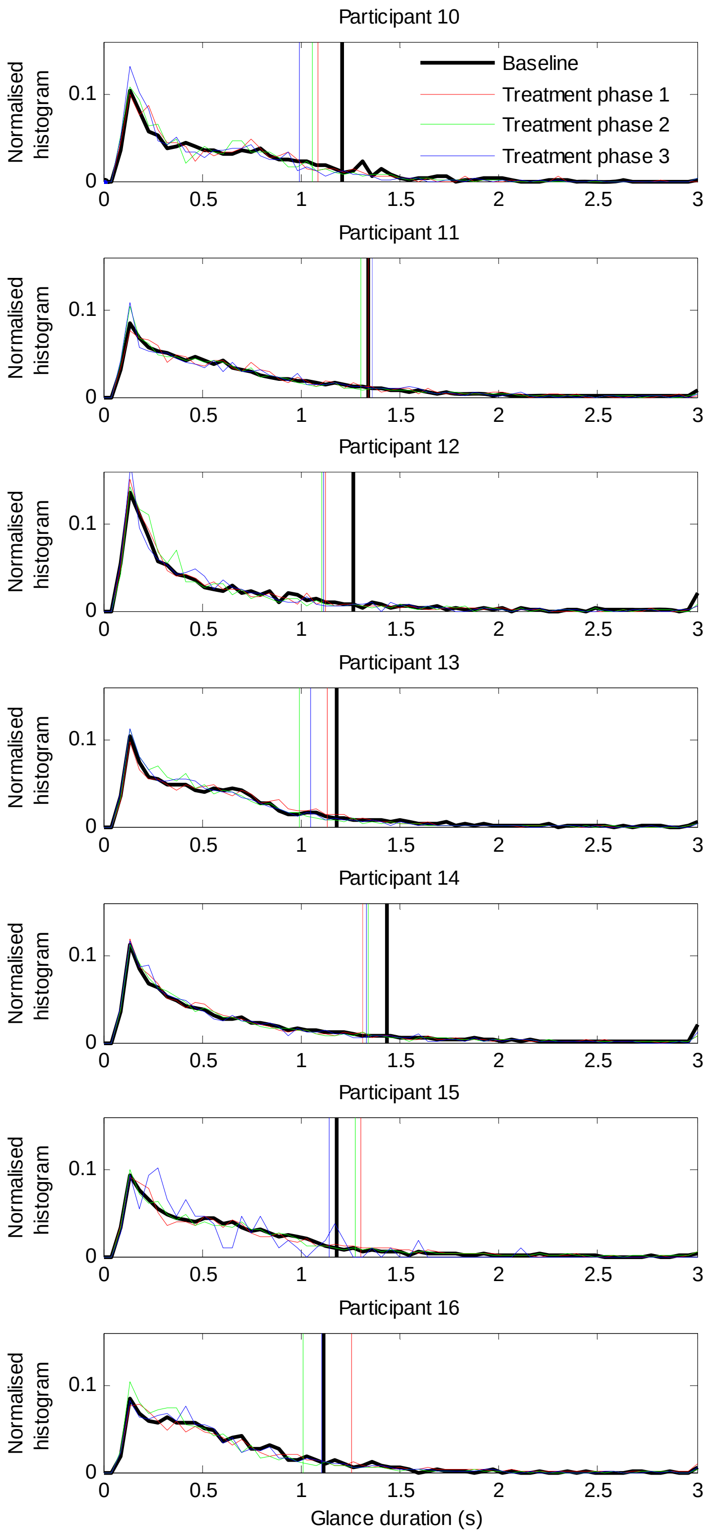


Figure 20 Histograms of glance duration per participant per phase, for gaze tracking only. The vertical lines represent the 95th percentile for glance duration for each respective phase.

A large percentage of glances away from the FRD was shorter than 0.5 s for all participants (Figure 20). The location of the 95th percentile of glance duration differs between participants and in most cases also within participants between phases. For Participant

10 and Participant 13 the 95th percentile of glance duration lies between 1 s and 1.3 s approximately, whereas it lies between 1.3 s and 1.5 s for Participant 14. Participant 11 exhibits practically no difference in the distributions for the different phases. For Participants 10, 12, 13 and 14 it appears that the 95th percentile of glance duration in the baseline phase is longer than in the treatment phases. Data are more scarcer for Participants 15 and 16 due to the frequent computer crashes, which makes their results less reliable.

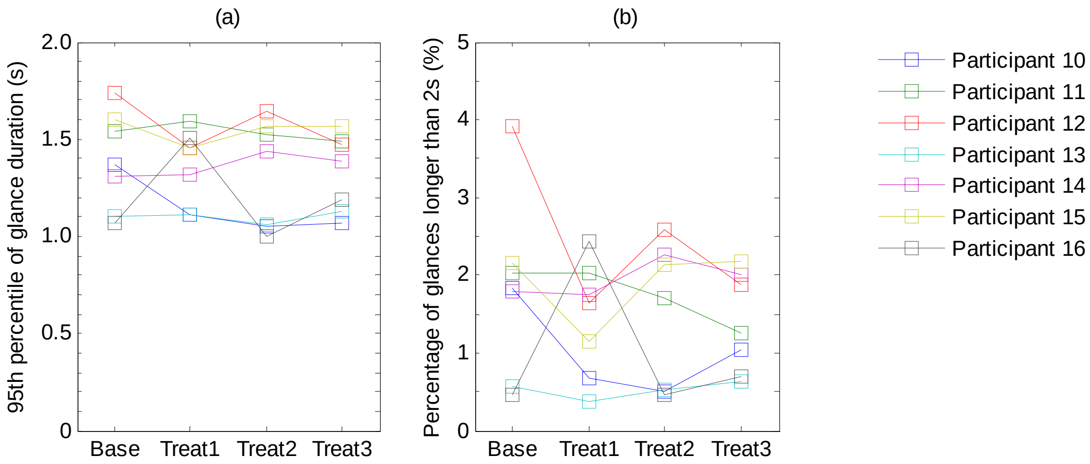


Figure 21 Long glance statistics per participant and phase for gaze tracking only. In (a) the 95th percentile of glance duration and in (b) the percentage of glances longer than 2 s is presented.

In Figure 21 the 95th percentile of glance duration is presented once more, side by side with the percentage of glances with a glance duration of more than 2 s. Those drivers whose 95th percentile of glance duration was higher also tended to have a higher percentage of glances above 2 s. Participant 10 exhibited the least amount of long glances with around 1%, for Participant 14 between 4 and 5% of all glances away from the FRD exceeded 2 s in duration. Except for Participants 15 and 16 both the 95th percentile and the percentage of glances above 2 s remained stable over the phases or had a tendency to decrease.

6.2 Discussion of glance duration

A comparison across participants shows that there are substantial differences between participants for the 95th percentile of glance duration, and also for the percentage of glances above 2 s. The 95th percentile has a range of about 0.5 s to 0.7 s, which is quite large, considering that the lowest value lies at about 1.35 s. Those participants who had higher values at the 95th percentile in general also had a larger percentage of glances above a duration of 2 s. When no inhibition is present, every glance longer than 2 s invariably leads to a distraction warning, given that tracking is available. Therefore a very consequent “classical conditioning” that those glances are inappropriate should take place, and a reduction of the occurrence of those long glances would reflect that drivers want to avoid the warning, and do not wait for it to occur and react then.

The glance duration data from gaze tracking suggest that the distraction warning system does not have a negative effect on glance duration, and for four of seven participants a positive effect could be observed. Participant 16 is the only one with a marked increase in glance duration in the first treatment phase, and it is likely that this effect can be explained by the technical problems encountered. Even though no inferential statistics were computed it appears that a reduction of the 95^{th} percentile of glance duration by about 0.1 to 0.2 s is an effect that should not be neglected. It is also important to see that the percentage of glances longer than 2 s tended to be smaller in the treatment phases, and this effect was more apparent for those participants who had relatively high percentages of very long glances in the baseline phase, namely Participant 14 and 12. The effect can be seen to some extent for Participants 11 and 15, too.

In general the glance duration distribution as measured by an automatic eye tracker and determined via the AttenD algorithm as opposed to a fixation based algorithm seems to produce results that are comparable to those published in the literature (see the chapter on glance duration in Kircher, K., 2007). However, it still appears desirable to compute fixation based glance data, which will be done in further analyses of the data.

The results on glance duration presented here indicate that there might be an effect of the distraction warning system on glance behaviour, which should be seen as an incentive to perform more detailed analyses of glance duration and distribution on a lower level than trip-based.

7 Number of glances

The number of glances leading up to a warning is of interest, because it gives some information about the strategy employed by the driver when looking away from the FRD. If drivers look away for two seconds or more they will surely get a warning already at the first glance, if no inhibition is active and if tracking is available. If they look away for a shorter duration, but several times in a row, it depends on the time that passes in between whether a warning will be issued or not. The accumulated duration of the glances away plus the latency of 0.1 s after every glance has to exceed the accumulated duration of the glances back to the FRD by 2 s for a warning to be issued.

In this chapter only those glances away from the FRD that in fact lead up to a warning are taken into account. In Figure 22 this would be the three glances shaded with green between minus 19 s and minus 15 s, whereas the other glances, shaded with red, are not counted, because they do not result in a buffer of zero seconds. The glance series marked in red are not counted into the glances leading up to the next warning, because the buffer becomes full again after them without reaching zero first.

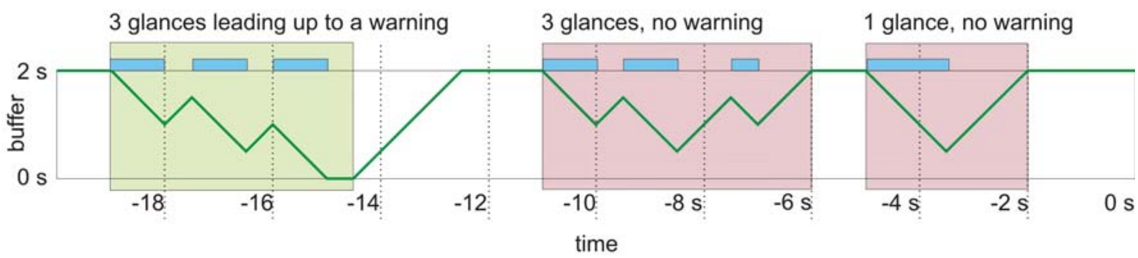


Figure 22 An example of a possible timeline of the attention buffer (green line). For simplification the 0.1 s latency is not included. The blue boxes indicate glances off the FRD.A green box indicates a succession of glances leading to a warning. A red box indicates a succession of glances not leading to a warning.

The statistics are presented for three types of situations in which the attention buffer reached zero. One subgroup are those occasions in which a warning was issued or had been issued if the condition had not been baseline, that is, all uninhibited warnings, both direct and indirect. Another subgroup is all inhibited warnings, regardless of what they were inhibited for. The third subgroup is all inhibited warnings that occurred at speeds higher than 50 km/h, which excludes all warnings that were inhibited due to speed. This group should resemble the uninhibited warnings more in terms of environment, as environments leading to a speed below 50 km/h often are qualitatively different from those environments in which higher speeds are allowed.

7.1 Glance count statistics

Besides glance duration and glance target, which were analysed in former chapters, the number of glances leading up to a warning was determined. The time span for counting glances started at the immediately preceding time instant when the attention buffer was full or at the immediately preceding warning, in case that the attention buffer had not filled up between two warnings. It stopped at the onset of the warning. In Figure 22 there is an example of a count from full buffer where three glances lead up to the warning. Both glances based on eye tracking and on nose tracking were taken into account.

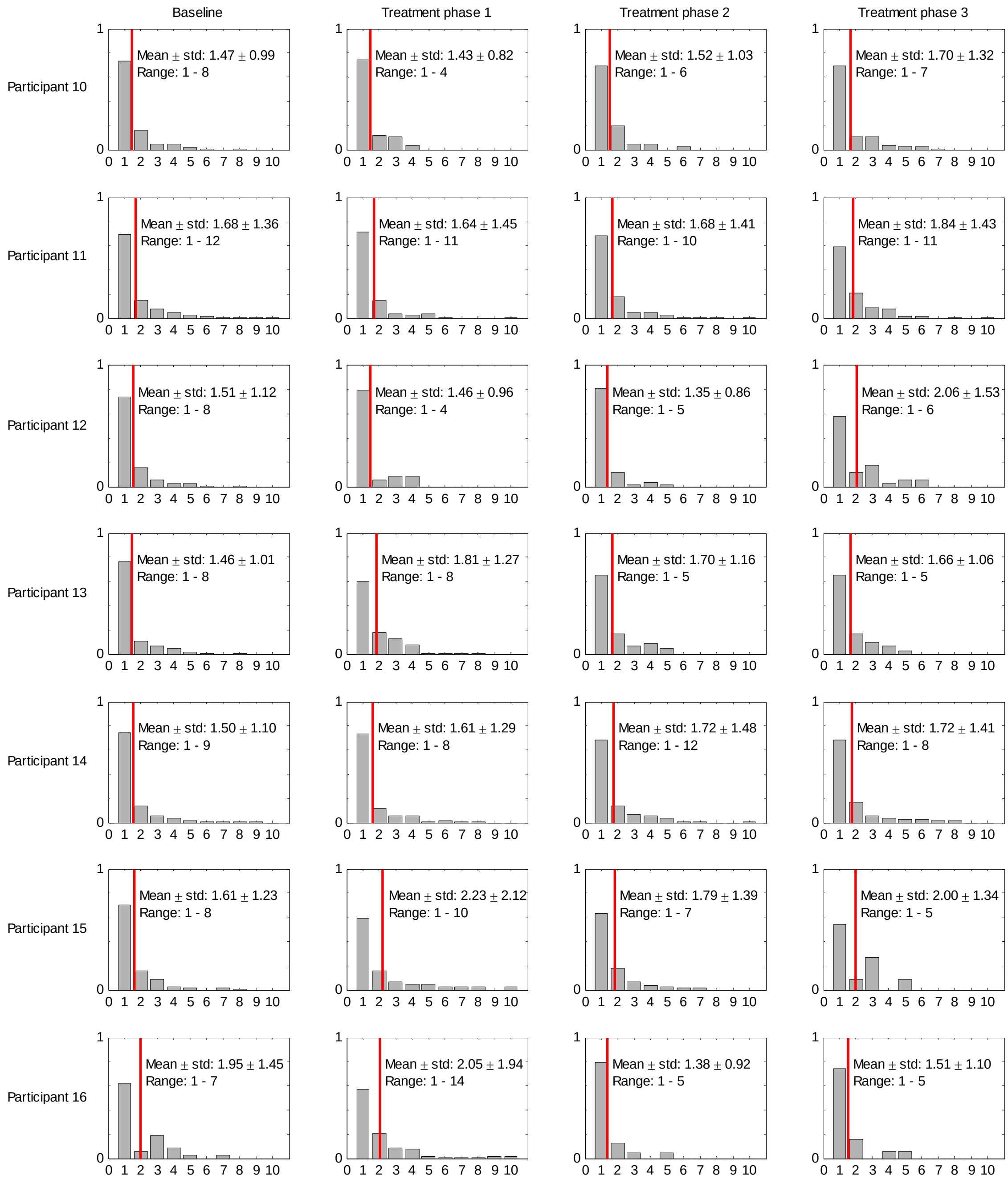


Figure 23 Bar graphs illustrating how many consecutive glances led to a warning per driver and phase (all glances add up to 1). The red line demarcates the mean for each driver and phase. Only warnings that ***were not inhibited*** *are considered.*

In Figure 23 the number of glances that led up to uninhibited warnings is presented. For all participants in most cases a single glance triggered a warning, but there were a number of occasions when the number of glances leading up to a warning was five or more. For those occasions when a single glance was enough, the glance had to be at least 2 s long.

The differences between participants are not big. Across participants, for 65 to 80 % of the warnings a single glance resulted in a warning. For the remaining 25 to 40 % of the warnings two or more glances led up to the warning, with a decreasing probability the higher the number of glances got. No clear trend could be found with respect to a change in the number of warnings between the baseline phase and the treatment phases.

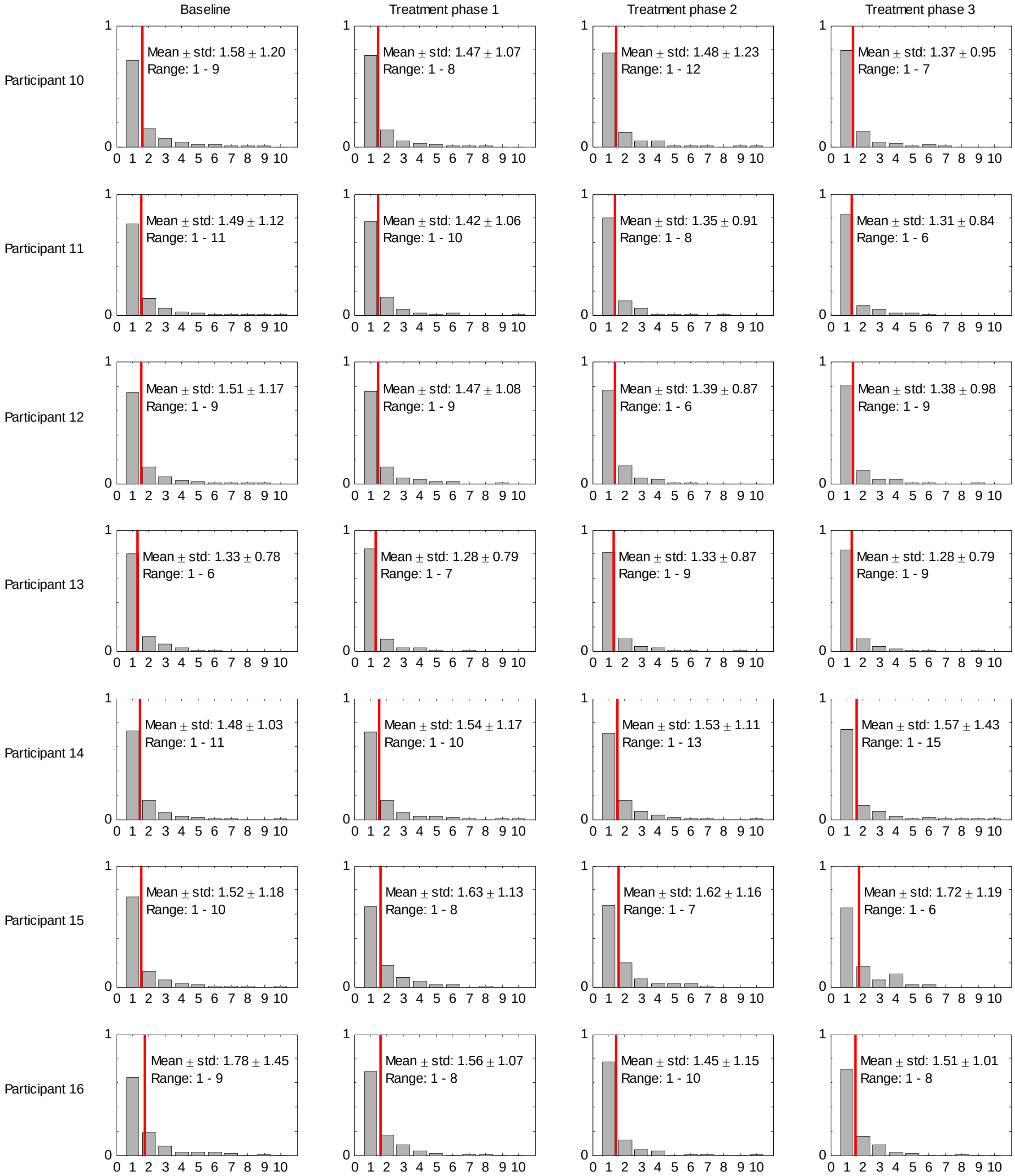


*Figure 24 Bar graphs illustrating how many consecutive glances led to a warning per driver and phase (all glances add up to 1). The red line demarcates the mean for each driver and phase. Only warnings that **were inhibited** are considered.*

In Figure24 the number of glances leading up to an inhibited warning is presented. In general the percentage of single glances is higher than for the uninhibited warnings. Both across participants and across phases no substantial differences could be found for the inhibited warnings.

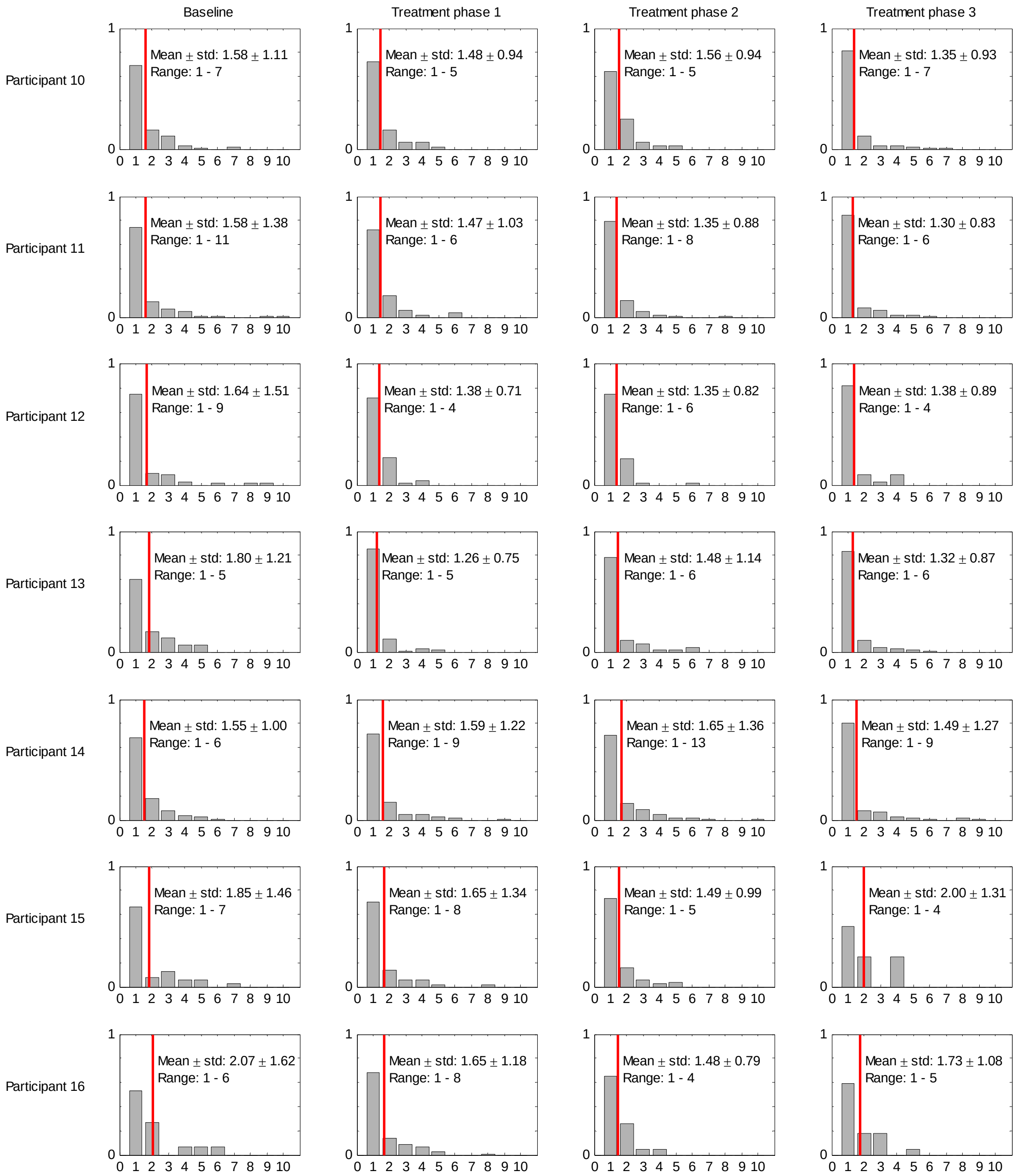


Figure 25 Bar graphs illustrating how many consecutive glances led to a warning per driver and phase (all glances add up to 1). The red line demarcates the mean for each driver and phase. Only warnings that ***were inhibited, and where the speed exceeded 50 km/h*** *are considered.*

The picture does not change much, either, when only glances leading up to inhibited warnings above 50 km/h are considered. By doing this, the warnings occurring due to the difference in speed are eliminated. Figure 25 and Figure 23 range across the same speeds, but split the glances into those leading up to an uninhibited versus an inhibited warning.

7.2 Discussion of the number of glances

It can be discussed whether it is positive or negative with only one glance leading up to a warning as compared to multiple glances. If only one glance results in a warning, the driver looks away from the FRD for at least two seconds, which is a relatively long consecutive time, but the total time on the secondary task up to the warning is shorter than for multiple glances. In the present analysis a new series of glances is always considered to begin after the buffer was filled up again. This implies that one glance or a series of glances after which the time during which the driver's gaze was directed at the FRD was long enough to fill the buffer completely was not included in the glance count, no matter how close in time they were to the starting point of the first counted glance. This is due to the fact that a driver is considered to be fully attentive when the attention buffer is full. It is thinkable, though, to come up with other criteria for when visual time sharing is supposed to begin and when it is supposed to end, which need not be based on the AttenD algorithm.

For about 25 % to 40 % of the uninhibited warnings the drivers had looked away from the FRD at least twice in a row without reaching full buffer in between. This indicates that in many cases of visual distraction the drivers still feel the urge to look back to the road relatively quickly, probably in order to confirm that they are still on the right track, and that nothing unexpected had come their way. Such confirmative glances back to the road are often very short, and it is not guaranteed that all necessary information can be processed, as was shown by Tsimhoni (2003) for example. Therefore, warnings are issued in those cases as well. It would be very informative to receive some feedback from the driver as to whether the warnings after single long glances or after repeated shorter glances feel more appropriate. It could be reasoned that a single long glance indicates that the driver "forgets about" the driving task and does not realise he is looking away from the road for such a long time. Therefore a warning might be considered not to be appropriate, because the own time judgement failed. If the driver still remembers looking back at the road and, thus, is more in a multitasking situation than completely distracted, the warning might feel more appropriate. The spontaneous comments given both during piloting and from some of the participants indicate such a relationship. Drivers felt that it was quite difficult to "provoke" a warning consciously, because they did not dare to look away from the road for so long. Warnings that came unprovoked were often not preceded by this feeling of press to look back to the road. Some drivers stated, however, that they realised only after having received the warning that they indeed had looked away for too long.

It had been planned to show drivers video recordings of certain situations during which they had received distraction warnings, in order to ask them about the appropriateness of the warnings and to get a more immediate understanding about the experienced effect of the warnings in particular situations. The relatively large number of warnings per participant prohibited this approach however, as it was felt that it would be difficult to recall one particular warning out of up to several hundred warnings. Still, this type of approach can be promising when single instances of warnings should be evaluated. For acceptance issues it would be of interest to find out whether there is a systematic difference in the experienced meaningfulness of warnings for single long glances as compared to multiple shorter glances.

It would also be of interest to investigate how much time passed between the last time the buffer was full until it was empty for the first time. This would reflect "time on task", summing up both when the driver looks away from the FRD and when he looks back, but has his mind presumably still on the secondary task. One approach could be to take the extraction method for gaze count used here as a definition for time on task, namely to start counting at the last time the buffer was full until it was empty. Additionally it appears meaningful to compute the time from full buffer to full buffer, and how many times it became zero in between. Furthermore, it should be considered how much time is allowed to pass between to glances away from the FRD, and whether preceding glances away could be included in the time on task, when they are close enough in time to the first glance included in the current definition.

8 Reaction time related behaviour

It is of interest to investigate whether the distraction warning makes the drivers look back to the FRD faster than they would without warning. If they did, this would be indicative of an immediate effect of the warning; prompting the driver to glance back to the FRD earlier than he would have otherwise. The AttenD algorithm used in this study considers a driver to be fully attentive when his attention buffer is full. Therefore it was also investigated how much time it took for the attention buffer to fill again after a warning, and how often the attention buffer reached zero after a warning before the buffer had been full again. If the time until the buffer is full again decreases for the treatment phases, this could be indicative of the driver's realising that he had been engaged in a secondary task, and that he more quickly dedicates his attention to the traffic again due to the warning. Likewise, if the attention buffer did not go down to zero as often during the treatment phases than during the baseline phase after a warning had been issued, the driver was probably prompted by the warning to give up prematurely on his secondary task, or to take a longer break from it than he would have otherwise.

Thus, in this chapter the time that passed from the onset of the warning until the driver looked back to the FRD, the time from the onset of the warning until the attention buffer was full again, and the number of times the attention buffer reached zero until it was full again are presented. It is differentiated between uninhibited warnings, inhibited warnings and inhibited warnings at speeds above 50 km/h, just like in Chapter 7.

In the tables below, the "time to FRD" always includes the 0.1 s latency period, meaning that the driver actually had glanced back to the FRD 100 ms faster than indicated in the tables. Just like in the other chapters, the two first days in each phase were excluded. Glances for which tracking was lost when the buffer equalled zero were excluded as well.

For the analyses both direct and indirect warnings were included, but warnings where the time until the attention buffer increased exceeded 3,000 ms were excluded, because this usually happened when tracking was lost while the attention buffer lay at zero. Warnings from trips shorter than 3 km/h were included in the analyses. Therefore, the number of warnings in each phase differs from the number of warnings reported in Chapter 4.

8.1 Participant 10

For Participant 10 for the uninhibited warnings no difference in "reaction time" could be found between the four phases (Table 29). The participant needed 0.8 s on average to look up to the FRD, meaning that 0.9 s passed before a count up was registered in the attention buffer. The time until the buffer was full was shortened significantly, however. The buffer was filled more than 2 s faster during the first treatment phase than during the baseline phase. Later on, however, the time until the buffer was filled increased again slightly, but did not reach the same level as during baseline driving. The number of times the buffer reached zero before it became full again decreased significantly from the baseline to the treatment condition, even though the same phenomenon as for the time to full buffer could be observed. After a sharp drop from baseline to the first treatment phase, a slight increase was found.

For the inhibited warnings, both in total and also for only those that occurred at speeds above 50 km/h, no changes could be found for any of the three indicators. The "reaction time" appears to be shorter for inhibited warnings above 50 km/h than for inhibited warnings in general and uninhibited warnings. The time that passed between a warning and the next time the buffer was full seems to be longer for inhibited warnings, and the buffer reached zero more often before it was full again.

Table 29 The time that passed after a warning until the buffer increased again, the time between a warning and full buffer, and the number of times the buffer turned 0 before it became full again for uninhibited warnings, inhibited warnings and inhibited warnings at speeds > 50 km/h per phase and in total.

		baseline	treat 1	treat 2	treat 3	total treat	total
uninhibited warning	time to buffer increase after warning (ms)	844 (n = 151)	1,087 (n = 28)	880 (n = 46)	947 (n = 149)	951	907 (n. s.)
	time to full buffer after warning (ms)	7,675 (n = 151)	5,566 (n = 28)	6,333 (n = 45)	6,455 (n = 149)	6,318	6,868 (p < .05)
	times buffer zero before full buffer after warning	1.37 (n = 151)	.86 (n = 28)	.91 (n = 45)	.99 (n = 149)	0.95	1.12 (p < .05)
inhibited warning	time to buffer increase after warning (ms)	925 (n = 1057)	938 (n = 453)	963 (n = 544)	913 (n = 845)	934	931 (n. s.)
	time to full buffer after warning (ms)	11,487 (n = 903)	12,741 (n = 389)	11,304 (n = 393)	12,676 (n = 778)	12,347	12,032 (n. s.)
	times buffer zero before full buffer after warning	2.11 (n = 905)	2.15 (n = 389)	1.63 (n = 394)	2.58 (n = 779)	2.23	2.19 (n. s.)
inhibited warning, speed > 50 km/h	time to buffer increase after warning (ms)	734 (n = 227)	629 (n = 35)	607 (n = 40)	689 (n = 163)	666	699 (n. s.)
	time to full buffer after warning (ms)	8,494 (n = 226)	5,130 (n = 35)	8,018 (n = 40)	11,528 (n = 163)	9,997	9,265 (n. s.)
	times buffer zero before full buffer after warning	1.61 (n = 226)	1.46 (n = 35)	1.28 (n = 40)	2.32 (n = 163)	2.02	1.82 (n. s.)

8.2 Participant 11

No significant changes in any of the three measures were observed for Participant 11 for the uninhibited warnings (Table 30). The data for the inhibited warnings are much less consistent, and a number of significant differences were found. For the inhibited warnings treatment phase 2 appears to have unusually short times until full buffer, and unusually few times for the buffer to reach zero before it became full. There is no clear pattern for the inhibited warnings above 50 km/h. It is noticeable, though, that the time until full buffer for uninhibited warnings is only about half to two thirds of the time to full buffer for inhibited warnings. However, this already holds true during the baseline phase, and the difference is smaller if only the inhibited warnings above 50 km/h are considered.

Table 30 The time that passed after a warning until the buffer increased again, the time between a warning and full buffer, and the number of times the buffer turned 0 before it became full again for uninhibited warnings, inhibited warnings and inhibited warnings at speeds > 50 km/h per phase and in total.

		baseline	treat 1	treat 2	treat 3	total treat	total
uninhibited warning	time to buffer increase after warning (ms)	768 (n = 581)	785 (n = 99)	681 (n = 191)	813 (n = 182)	754	762 (n. s.)
	time to full buffer after warning (ms)	7,782 (n = 581)	7,830 (n = 99)	6,353 (n = 191)	8,946 (n = 182)	7,663	7,728 (n. s.)
	times buffer zero before full buffer after warning	1.55 (n = 581)	1.17 (n = 99)	.96 (n = 191)	2.04 (n = 182)	1.42	1.50 (n. s.)
inhibited warning	time to buffer increase after warning (ms)	920 (n = 2237)	936 (n = 388)	934 (n = 588)	946 (n = 747)	940	928 (n. s.)
	time to full buffer after warning (ms)	12,546 (n = 2182)	11,202 (n = 355)	8,894 (n = 546)	12,273 (n = 717)	10,898	11,844 (p < .05)
	times buffer zero before full buffer after warning	2.83 (n = 2182)	2.05 (n = 356)	1.65 (546)	2.81 (n = 717)	2.25	2.59 (p < .05)
inhibited warning, speed > 50 km/h	time to buffer increase after warning (ms)	780 (n = 1024)	796 (n = 132)	798 (n = 239)	880 (n = 359)	838 (n = 730)	805 (p < .05)
	time to full buffer after warning (ms)	11,422 (n = 1024)	8,341 (n = 132)	8,717 (n = 239)	11,305 (n = 359)	9,922 (n = 730)	10,797 (p < .05)
	times buffer zero before full buffer after warning	3.21 (n = 1024)	1.54 (n = 132)	1.88 (n = 239)	3.18 (n = 423)	2.46 (n = 730)	2.90 (p < .05)

8.3 Participant 12

For Participant 12 no differences in reaction time between the baseline phase and the treatment phases were found both for uninhibited warnings (Table 31). On average it took around 900 ms for the driver to look at the road again, with the 100 ms latency period subtracted from the time to FRD. The time to full buffer was somewhat shorter during the treatment phases than during baseline, but the difference was not statistically significant. However, there was a clear and significant reduction of the number of times the buffer reached zero before becoming full again from the baseline phase to the treatment phases.

When looking at all inhibited warnings, a similar pattern turns up. The reaction time is about equal in length as for the uninhibited warnings, but it takes longer for the buffer to fill up again. The times at zero after a warning decrease significantly for the treatment phase, even though the warnings actually were inhibited. If only the inhibited warnings above 50 km/h are considered, the effect is no longer statistically significant, even though it is quite clear at least for treatment phases one and three. There is no immediate explanation why the value in treatment phase 2 is close to baseline levels.

Table 31 The time that passed after a warning until the buffer increased again, the time between a warning and full buffer, and the number of times the buffer turned 0 before it became full again for uninhibited warnings, inhibited warnings and inhibited warnings at speeds > 50 km/h per phase and in total.

		baseline	treat 1	treat 2	treat 3	total treat	total
uninhibited warning	time to buffer increase after warning (ms)	938 (n = 77)	1,127 (n = 45)	1,037 (n = 51)	983 (n = 56)	1,044	1,008 (n. s.)
	time to full buffer after warning (ms)	9,546 (n = 77)	8,438 (n = 45)	8,556 (n = 51)	7,908 (n = 56)	8,283	8,707 (n. s.)
	times buffer zero before full buffer after warning	2.01 (n = 77)	1.33 (n = 45)	1.63 (n =51)	.96 (n = 56)	1.30	1.54 (p = .05)
inhibited warning	time to buffer increase after warning (ms)	1,008 (n = 930)	968 (n = 460)	1,018 (n = 424)	1,038 (n = 424)	1,007	1,007 (n. s.)
	time to full buffer after warning (ms)	12,194 (n = 888)	13,092 (n = 413)	9,153 (n = 404)	11,724 (n = 413)	11,339	11,697 (n. s.)
	times buffer zero before full buffer after warning	2.23 (n = 891)	2.04 (n = 414)	1.63 (n = 404)	2.14 (n = 413)	1.94	2.06 (p < .05)
inhibited warning, speed > 50 km/h	time to buffer increase after warning (ms)	915 (n = 204)	835 (n = 82)	759 (n = 92)	818 (n = 74)	802	953 (n. s.)
	time to full buffer after warning (ms)	9,808 (n = 204)	8,931 (n = 82)	8,460 (n = 92)	9,811 (n = 74)	9,019	9,375 (n. s.)
	times buffer zero before full buffer after warning	2.27 (n = 204)	1.51 (n = 82)	2.14 (n = 92)	1.65 (n = 74)	1.79	2.00 (n. s.)

8.4 Participant 13

For this participant, just as for the others, no changes in reaction time to the uninhibited warnings could be found. The time to full buffer changed significantly over the phases, with the treatment phase 1 having the longest time to full buffer and treatment phase 2 having the shortest. Even though the number of times the buffer became zero before filling up changed substantially across phases, no statistically significant changes could be measured.

Generally, the data vary quite substantially across phases for this participant. The number of warnings per phase and warning criterion is not so low, however, that these changes could be attributed to chance or idiosynchratic circumstances.

Table 32 The time that passed after a warning until the buffer increased again, the time between a warning and full buffer, and the number of times the buffer turned 0 before it became full again for uninhibited warnings, inhibited warnings and inhibited warnings at speeds > 50 km/h per phase and in total.

		baseline	treat 1	treat 2	treat 3	total treat	total
uninhibited warning	time to buffer increase after warning (ms)	862 (n = 136)	887 (n = 131)	864 (n = 100)	847 (n = 104)	867 (n = 335)	866 (n. s.)
	time to full buffer after warning (ms)	7,523 (n = 136)	9,834 (n = 131)	6,077 (n = 100)	7,721 (n = 104)	8,057 (n = 335)	7,902 ($p < .05$)
	times buffer zero before full buffer after warning	1.65 (n = 136)	2.12 (n = 131)	.91 (n = 100)	1.69 (n = 104)	1.63 (n = 335)	1.63 (n. s.)
inhibited warning	time to buffer increase after warning (ms)	998 (n = 879)	1,022 (n = 748)	1,030 (n = 793)	1,074 (n = 737)	1,041 (n = 2278)	1,030 (n. s.)
	time to full buffer after warning (ms)	13,333 (n = 802)	15,103 (n = 698)	12,396 (n = 726)	14,981 (n = 689)	14,133 (n = 2113)	13,913 (n. s.)
	times buffer zero before full buffer after warning	2.16 (n = 802)	3.54 (n = 699)	1.92 (n = 726)	3.04 (n = 690)	2.82 (n = 2115)	2.64 ($p < .05$)
inhibited warning, speed > 50 km/h	time to buffer increase after warning (ms)	829 (n = 209)	1,041 (n = 247)	741 (n = 104)	888 (n = 178)	930 (n = 529)	902 ($p < .05$)
	time to full buffer after warning (ms)	9,166 (n = 209)	20,942 (n = 247)	7,261 (n = 104)	15,946 (n = 178)	16,571 (n = 529)	14,474 ($p < .05$)
	times buffer zero before full buffer after warning	2.61 (n = 209)	6.57 (n = 247)	.95 (n = 104)	6.26 (n = 178)	5.36 (n = 529)	4.58 ($p < .05$)

8.5 Participant 14

Again, no effect of the warning system on reaction time could be found for the uninhibited warnings. Otherwise, the analyses show somewhat unexpected results for this participant. The time to full buffer increases significantly when the warning system comes on, both for uninhibited and inhibited warnings above 50 km/h. Additionally, the buffer reached zero between a warning and full buffer significantly more times with the warning system active than during baseline. The number of warnings is exceptionally large for each phase for this participant, which is related to the huge mileage that he collected during the course of the month. This high number of warnings makes it very unlikely that the results are artefacts.

Table 33 The time that passed after a warning until the buffer increased again, the time between a warning and full buffer, and the number of times the buffer turned 0 before it became full again for uninhibited warnings, inhibited warnings and inhibited warnings at speeds > 50 km/h per phase and in total.

		baseline	treat 1	treat 2	treat 3	total treat	total
uninhibited warning	time to buffer increase after warning (ms)	1,052 (n = 816)	1,088 (n = 125)	1,070 (n = 274)	1,047 (n = 242)	1,065 (n = 641)	1,058 (n. s.)
	time to full buffer after warning (ms)	9,598 (n = 816)	9,849 (n = 125)	9,686 (n = 274)	11,747 (n = 242)	10,496 (n = 641)	9,993 (p < .05)
	times buffer zero before full buffer after warning	1.42 (n = 816)	1.26 (n = 125)	1.41 (n = 274)	2.31 (n = 242)	1.72 (n = 641)	1.55 (p < .05)
inhibited warning	time to buffer increase after warning (ms)	961 (n = 2682)	961 (n = 652)	953 (n = 1186)	1,022 (n = 1445)	985 (n = 3283)	974 (p < .05)
	time to full buffer after warning (ms)	11,268 (n = 2576)	10,569 (n = 629)	11,291 (n = 1099)	11,657 (n = 1337)	11,302 (n = 3065)	11,287 (n. s.)
	times buffer zero before full buffer after warning	1.77 (n = 2576)	1.91 (n = 630)	2.11 (n = 1100)	2.43 (n = 1339)	2.21 (n = 3069)	2.01 (p < .05)
inhibited warning, speed > 50 km/h	time to buffer increase after warning (ms)	902 (n = 1394)	997 (n = 167)	915 (n = 496)	925 (n = 622)	930 (n = 1285)	916 (n. s.)
	time to full buffer after warning (ms)	10,793 (n = 1394)	12,508 (n = 167)	13,985 (n = 496)	13,780 (n = 620)	13,694 (n = 1283)	12,183 (p < .05)
	times buffer zero before full buffer after warning	2.00 (n = 1394)	2.69 (n = 167)	2.96 (n = 496)	3.56 (n = 620)	3.22 (n = 1283)	2.58 (p < .05)

8.6 Participant 15

For Participant 15 a significant increase in reaction time could be observed over time for uninhibited warnings. The time until full buffer varied without a clear direction, as did the number of times the buffer was zero before it became full again. Similar not very consistent results were found for the inhibited warnings. Especially treatment phase 3 can be viewed with some caution in this case, because only 18 uninhibited warnings were registered, due to the faulty log computer.

Table 34 The time that passed after a warning until the buffer increased again, the time between a warning and full buffer, and the number of times the buffer turned 0 before it became full again for uninhibited warnings, inhibited warnings and inhibited warnings at speeds > 50 km/h per phase and in total.

		baseline	treat 1	treat 2	treat 3	total treat	total
uninhibited warning	time to buffer increase after warning (ms)	685 (n = 116)	855 (n = 38)	931 (n = 111)	1,273 (n = 18)	950 (n = 167)	842 p < .05)
	time to full buffer after warning (ms)	7,278 (n = 116)	8,955 (n = 38)	6,218 (n = 111)	7,245 (n = 18)	6,951 (n = 167)	7,085 (p < .05)
	times buffer zero before full buffer after warning	1.22 (n = 116)	1.53 (n = 39)	.82 (n = 111)	1.22 (n = 18)	1.02 (n = 167)	1.11 (n. s.)
inhibited warning	time to buffer increase after warning (ms)	901 (n = 754)	960 (n = 512)	912 (n = 554)	947 (n = 140)	936 (n = 1206)	923 (n. s.)
	time to full buffer after warning (ms)	10,067 (n = 703)	10,846 (n = 458)	9,195 (n = 512)	8,355 (n = 96)	9,828 (n = 1066)	9,923 (n. s.)
	times buffer zero before full buffer after warning	1.77 (n = 703)	2.15 (n = 458)	1.63 (n = 513)	1.59 (n = 96)	1.85 (n = 1067)	1.82 (p < .05)
inhibited warning, speed > 50 km/h	time to buffer increase after warning (ms)	804 (n = 162)	876 (n = 78)	966 (n = 159)	801 (n = 34)	919 (n = 271)	876 (n. s.)
	time to full buffer after warning (ms)	7,306 (n = 162)	7,311 (n = 68)	9,260 (n = 150)	6,485 (n = 25)	8,429 (n = 243)	7,980 (n. s.)
	times buffer zero before full buffer after warning	1.11 (n = 162)	1.15 (n = 68)	1.79 (n = 151)	1.32 (n = 25)	1.56 (n = 244)	1.38 (p = .05)

8.7 Participant 16

For this participant the “reaction time” was higher during the treatment phases as compared to baseline driving, even though the difference was not statistically significant. Both the time until full buffer and the number of times the buffer reached 0 before it reached 2 s increased from baseline to the first treatment phase, but afterwards they decreased below baseline level. The significant difference in the time to full buffer was only between the first and the second treatment phase. For the inhibited warnings significant effects were only found for the number of times the buffer reached zero, and there were relatively large and not easily predictable jumps over the different periods.

Table 35 The time that passed after a warning until the buffer increased again, the time between a warning and full buffer, and the number of times the buffer turned 0 before it became full again for uninhibited warnings, inhibited warnings and inhibited warnings at speeds > 50 km/h per phase and in total.

		baseline	treat 1	treat 2	treat 3	total treat	total
uninhibited warning	time to buffer increase after warning (ms)	649 (n = 30)	847 (n = 154)	761 (n = 52)	988 (n = 57)	860 (n = 263)	839 (n. s.)
	time to full buffer after warning (ms)	6,177 (n = 30)	7,515 (n = 153)	5,065 (n = 52)	5,636 (n = 57)	6,620 (n = 292)	7,085 (p < .05)
	times buffer zero before full buffer after warning	1.03 (n = 30)	1.20 (n = 153)	.56 (n = 52)	.75 (n = 57)	0.97 (n = 262)	0.98 (n. s.)
inhibited warning	time to buffer increase after warning (ms)	918 (n = 254)	897 (n = 464)	849 (n = 276)	871 (n = 298)	877 (n = 1038)	885 (n. s.)
	time to full buffer after warning (ms)	10,543 (n = 222)	9,029 (n = 457)	10,539 (n = 260)	10,943 (n = 283)	9,963 (n = 1000)	10,069 (n. s.)
	times buffer zero before full buffer after warning	1.99 (n = 223)	1.65 (n = 457)	1.80 (n = 260)	1.30 (n = 283)	1.59 (n = 1000)	1.66 (p < .05)
inhibited warning, speed > 50 km/h	time to buffer increase after warning (ms)	772 (n = 71)	873 (n = 245)	795 (n = 66)	913 (n = 57)	865 (n = 368)	850 (n. s.)
	time to full buffer after warning (ms)	8,389 (n = 59)	10,178 (n = 244)	9,577 (n = 63)	7,079 (n = 54)	9,609 (n = 361)	9,438 (n. s.)
	times buffer zero before full buffer after warning	1.75 (n = 59)	2.00 (n = 244)	2.05 (n = 63)	.98 (n = 54)	1.85 (n = 361)	1.84 (p < .05)

8.8 Comparison between participants

On average the time from the onset of the (silent) warning until the attention buffer increased again varied between 0.6 and 1.0 s between participants for the baseline phase. Averaged across the treatment phases the time from the onset of the warning until the attention buffer increased lay between 0.7 and 1.1 s across participants. This shows that the participants did not look up to the FRD more quickly after having received a warning than during baseline driving.

It took between 6 and 9 seconds until the buffer was full again after an uninhibited (silent) warning in the baseline phase, averaged across the treatment phases it took from between 6 and 10 seconds for the buffer to reach 2 s. For two participants the average time until full buffer decreased significantly from baseline to average treatment, for three participants it increased significantly, and for 2 participants there was no significant difference between the baseline and the accumulated treatment phase. The time to full buffer varied between weeks within subjects, however.

On average the attention buffer becomes empty at least once more after a warning before full buffer is reached. For two participants this value decreased significantly between baseline and treatment, for one participant it increased significantly, while for the remaining drivers no statistical difference could be found.

The time from the onset of an inhibited warning until the buffer increased again is approximately equal to that found for the uninhibited warnings. The time until the buffer is filled up to 2 s again after having been completely depleted is longer in the case of inhibited warnings. This is also reflected by the fact that the buffer reaches 0 more often before becoming full again for inhibited than for uninhibited warnings. This number varies between participants from about once to more than 5 times on average for inhibited warnings over baseline and treatment together.

8.9 Discussion of reaction time

The distraction warning system in its present form does not have the direct effect of shortening the duration of the last glance away from the FRD. Except for Participant 15 the reaction time did not change significantly between the baseline and the three treatment phases, and for Participant 15 the reaction time increased substantially over the phases. This is not easy to explain and might be an artefact due to the frequent computer crashes. In general it takes on average between 0.6 and 1.0 s from the onset of an uninhibited warning until the attention buffer starts to increase again. The actual reaction time from the onset of the warning until the gaze direction is registered to be directed back at the FRD is 0.1 s shorter than the value reported in the tables, however, due to the latency period of the AttenD algorithm.

There is no consistent picture for the average time that passed since an uninhibited warning was issued until the buffer became full again. Generally the time until full buffer after an uninhibited warning lies at about six to ten seconds, which is markedly more than the minimum time of two seconds plus the reaction time. This implies that drivers do not get prompted by the warning to look back at the FRD immediately and keep the glance there for at least two seconds, but rather that the drivers look away from the FRD again before the buffer filled up, or that the system loses tracking before the buffer is full. Loss of tracking is often an indication for the driver's looking away from the FRD.

A further indication that drivers do not look back to the FRD long enough for the buffer to fill up completely is the number of times the buffer turns zero before filling up. On average this happens around 0.5 to 2 times for uninhibited warnings. This behaviour can be interpreted such that there appear to be many instances when drivers basically ignore the warning and keep on looking at what they had been looking at before the warning came.

No difference in reaction time could be found between inhibited and uninhibited warnings, which are a further indication that the warning in itself did not speed up the drivers' looking back to the FRD. For all drivers, however, the time that passed until the buffer was full again was on average several seconds longer for inhibited warnings than for uninhibited warnings. In most cases the average duration for the inhibited warnings above 50 km/h lay somewhere in the middle between the uninhibited warnings and all inhibited warnings, indicating that it took the longest time for the buffer to fill up at speeds below 50 km/h. This indicates that especially at those low speeds a lot of time is spent scanning the environment at a much wider angle than at higher speeds.

The average number of times with which the buffer reaches zero before it fills up completely is in general somewhat higher for inhibited warnings than for uninhibited warnings, but here no marked difference between all inhibited warnings and only those above 50 km/h can be found. This could be an indication that drivers look around more while driving at lower speeds, but that those glances usually are short enough in order to avoid emptying the buffer completely.

A much more thorough analysis of the glance behaviour in relation to occurrences on the road and the roadside, and how the glance behaviour looked like both before and after a warning would be needed to be able to make more definite statements on the effects of the distraction warning on glance behaviour in immediate relation to the warning.

9 Self-reported results

Among all data collected during the field study a number of questionnaires were filled in by the participants. During the first part of the driving session the drowsiness and distraction detection system was in baseline mode, only collecting data, but no interaction with the driver was present. This means that no warnings were given to drivers, and the car was just as any normal car. After this time the second questionnaire was filled in. Then the warning system was set to enable warnings. The drivers completed another three weeks of driving with the system active and warning the driver in case of detected drowsiness or distraction, and finally the third questionnaire was filled in. Since all drivers were Swedes, the questionnaires were in Swedish language.

All participants agreed to be part of the study and allowed that data would be collected, analysed and presented. Here each participant is only identified by a number and the gender.

Data from seven subjects were available, this number is too low to allow statistical inference, thus the data analysis focussed on qualitative aspects. Results are presented and discussed at the same time. A total of 12 questions were answered before the drive, 26 questions in-between the drives, and 31 questions after the drive. Many questions contained several sub-questions, the total number of sub-questions is 226. The response rate was very high. The questionnaires are included in the appendix of the technical methods report (Kircher, K., Kircher, A. & Claezon, F., 2009).

Data were analysed with SPSS 15.0.

9.1 Questionnaires

The questionnaires were filled in by the participants before starting the drive (named “before”), after the baseline (named “between”), and after finishing the drive and having experienced the system with warnings active (named “after”). All results presented here are self-reported. This means that they do not necessarily have to correspond completely to what was found in the logged data.

9.1.1 General results

Four male and three female participants completed the drive. Their mean age was 42 years (sd = 10.9 years), but observe that one of the participants did not report his age. They had held their driving license for 25 years (± 10.9 years) on average. Their average mileage in the last year was ca. 36,000 km (± 14,840 km), and their estimated total mileage was around 500,000 km (±242,585 km). The drivers had very little experience of ITS in their private or company cars: Only two drivers had used navigation systems or parking assistance or park distance warning systems before. None had used any advanced driving assistance systems.

Six of the seven participants use to drive the same routes often or very often, for example from home to work and back. These trips were most often driven between 6 AM and 9 PM. Seen across the whole group, trips in urban, rural and highway environment were quite evenly distributed.

9.1.2 Results before driving

Four of the seven drivers assessed their driving style as neutral, while one driver rated herself as somewhat defensive and two rated themselves as somewhat offensive. Most participants believed that driving long trips without break would increase drowsiness, while two answered that their drowsiness level would be the same after a long trip. Four participants assumed that reaction time would degrade after having driven for a long time, while three answered there would be no change in reaction time.

9.1.3 Results after baseline driving with the system deactivated (“between”)

After having driven the experimental car for approximately ten days the participants rated that they were driving more or less as they did with their private car with respect to driving style, frequency and the ”quality of driving”. Driving pleasure had increased for five of the seven drivers. This might be a result of the car being relatively fancy, and in most cases above the standard of the participants’ own cars. Getting adapted to the new car took only one day for four of the drivers, while all were adapted within 4 days.

With regard to potentially hazardous behaviour all participants believed driving under the influence of alcohol was inacceptable, but only one driver thought driving using a hands-free mobile phone was “somewhat inacceptable”. Driving while drowsy was seen as inacceptable by all but one driver.

Opinions about which ITS would increase traffic safety varied greatly; for distraction warning systems the opinion was mostly positive, for the drowsiness warning system it was only positive, while for navigation systems, collision warning systems, lane departure warning system, and slippery road detection systems negative opinions about the traffic safety effects could be found. Being inattentive for a short period during driving while performing other tasks was considered dangerous by all drivers in urban environments, but on roads with speed limits from 70 to 110 km/h the opinion varied greatly between safe and dangerous. Among possibly distracting activities while driving using a handheld mobile phone was most common. Interestingly enough, using a hand-held mobile phone was also seen as the most distracting non-driving related activity. Driving in drowsy or distracted condition happened rarely to most participants, but one driver reported driving often in very drowsy condition, often being distracted, and often having passengers pointing out hazardous situations. For just this driver, who could have benefitted very much from the drowsiness and distraction warning system, unfortunately the data collection was suboptimal, leading to significant data loss. Risk of experiencing drowsiness seemed to increase with the speed limit on the road: For urban roads all participants thought the risk was small, while on motorways it was large.

9.1.4 Expectations of the distraction warning system (“between”)

Before actually trying the distraction warning system all participants but one had a very positive general opinion. This person had a very negative opinion about the system. Just this participant had reported before to drive when very drowsy and distracted. Among all participants, this participant also reported the lowest expected benefit for the system, showing a coherent answering pattern.

The preferred warning modality for the system was a sound alert, while belt vibration received the lowest score. One question was related to possible risks the drivers could see with the distraction warning system. Five of the seven drivers could see a risk: Both over-trust in the system and additional distraction from the system were named, besides one driver expressing concern about the technical reliability of the system. It is therefore interesting that all drivers wrote they would trust the system “very much”. Both the distraction and the drowsiness warning systems received the same high level of trust from the drivers, and the foreseen risks were similar, too.

9.1.5 Expectations of the drowsiness warning system (“between”)

The eight questions related to the drowsiness warnings system were very similar to those about the distraction warning system. The general opinion about the drowsiness warning system was positive, and drivers thought they would benefit from the system. Again, the preferred warning modality was sound, with belt vibrations obtaining least preference. As mentioned, foreseen risks and level of trust were the same as for the distraction warning system, and in general the opinion about both systems was very similar.

9.1.6 Results after the test period (“after”)

These questionnaires were filled in after completing the entire test, which was after having had the car for approximately one month. The warning systems had been active during the last two thirds of the test period.

The first question was related to how safe drivers thought being inattentive for a short moment had been in different traffic environments. The results after having used the system pointed clearly towards lower risk awareness, that is, for all environments the drivers answered that it was less dangerous to be inattentive than they had said before using the system. This is interesting and requires further discussion.

The frequency of possibly distracting activities, such as using mobile phones, was reported to be approximately the same as before using the system. The same was valid for the frequency of possibly risky behaviour, like being very tired when driving. The distraction and drowsiness warnings were generally not perceived as being distracting, however, one driver answered that the distraction warnings were “very much” distracting.

9.1.7 Experience with the distraction warning system

All seven drivers reported having received several warnings from the system (20 to 100 warnings) and two drivers reported having received more than 100 warnings. After the test period the general opinion towards the distraction warning system was very positive; it received approximately the same ratings as before the system had been used. Five of the seven drivers thought the warnings rarely felt unnecessary and disturbing, while the other two drivers answered the warnings “often” felt unnecessary and disturbed “sometimes”. Both drivers who thought the warnings were often unnecessary and disturbed had received more than 100 warnings (self rated). Here it would be interesting to see if the system did not function correctly in the situations rated as disturbing, which would explain why the warnings felt this way. It is interesting to note that five drivers trusted the system completely, while two expressed a neutral opinion. One of these two drivers had received more than 100 warnings. The warning modality (vibration) did not receive major comments.

Traffic safety was experienced to be increased somewhat by the system, but the expectation score had been higher before using the system. All drivers answered they would like to have such a system in their car, and on average they would be ready to pay 8,000 SEK (approximately 800 Euro) for it.

9.1.8 Experience with the drowsiness warning system

Here the number of answers was lower than for the distraction warning system, since not all drivers had received drowsiness warnings: Four drivers had received one drowsiness warning (low level of drowsiness), and one had received many warnings (again low level), but no drivers had received warnings on the middle or high level of drowsiness.

The general opinion about the system was very positive, besides one driver answering "very negative" (this is the same driver who had expressed a very negative opinion about the systems before starting to use them). The warnings did not disturb the driver, but the rating for how often they had felt unnecessary varied from "very often" to "never". Note that only four drivers had received warnings. Six drivers wanted to have such a system in their car, even those who had not received any warnings, and thus could not know how the system performed. One driver did not answer the question. The drivers would be willing to pay 7,750 SEK (approximately 770 Euro) on average for the drowsiness warning system.

9.2 Interviews

Informal interviews were conducted with all drivers when they returned the vehicle. All but one expressed a very positive attitude towards the distraction warning system. They unanimously reported that the warnings had raised their awareness of what they were doing while driving a car. One participant even claimed to have changed his behaviour substantially. He would now refrain from telephone conversations while driving and stop the car when eating a sandwich. Some participants reported that they thought they might have received false warnings when they squinted against the sun.

9.3 Discussion of self-reported results

As mentioned before, the limited number of participants does not allow any statistical inference, but only qualitative observations. In general the experience with the distraction and drowsiness warning systems was positive for all drivers except one, and drivers would like to have such a system in their car. In general the answer pattern was quite homogenous, but one driver stood out with a very critical view of the systems. The chosen warning modality and timing of the warnings was not commented on negatively.

It is striking that most drivers report to have received less than 100 warnings during their time with the vehicle. All participants except for one, for whom the computer had become unstable, had received more than 100 warnings during the treatment phase, one of the drivers who had reported to have received between 20 and 100 warnings had, in fact, received more than 700 warnings. The only driver with less than 100 warnings had actually estimated to have received more than that. Most drivers substantially underestimate the number of warnings, which could be a sign for high acceptance. Warnings that are not experienced as very disturbing might not be remembered as much as annoying warnings. This argumentation is supported by the fact that the one driver who overestimated the number of received warnings had the most negative attitude towards the warning system.

Theoretically it could also be the case that the drivers did not feel the warnings at all, which would be a natural explanation for not remembering them. This appears to be very unlikely, however, because it rarely happens during driving above 50 km/h that the contact with the seat cushion is lost. When the system was demonstrated to the drivers before they set off for the treatment phase, no driver mentioned any difficulties in feeling the vibration. The vibration was very distinct and could not have been confused with vibrations resulting from a rough road.

It can be speculated that the drivers' sinking awareness of the risks associated with inattention is caused by the presence of the distraction warning system. The fact that many drivers report a high level of trust for the system makes this assumption even more valid. During informal interviews, however, drivers had reported that they felt that the system had increased their awareness of what they were doing while driving a car. The glance duration data do not give rise to the suspicion that drivers start abusing the system by "waiting out" a warning. "Trusting" a system does not necessarily mean that one wants to abuse the system, but rather, that it might reliably help in situations where the driver does not perform optimally. Therefore, high system trust in itself is not an immediate ground for suspicion. The data are somewhat contradictory here, and it is not immediately conclusive why inattention is seen as less risky after having driven with the system.

It has to be remembered, of course, that only seven drivers participated in the study, and they are not representative for the driving population at large. Still, the overall picture seems to be that a distraction warning system is desirable for drivers, and that they feel that they could benefit from it.

10 Difficulties encountered

A great number of the technical difficulties that were encountered were already addressed in the technical report on the equipment and method (Kircher, K., Kircher, A. & Claezon, F., 2009). In this report difficulties related to the analysis will be taken up.

As mentioned in the methods report, a bug in the AttenD algorithm could be fixed first after Participant 10 had driven the baseline phase. This led to some manual work filtering out the files that were affected by the bug. It appears like there was no systematic affliction of files by the bug, which should mean that the results are not slanted due to the removal of corrupted files. However, it took time to manually remove the affected files, which in turn affected the time remaining for data analyses.

Similar problems were encountered for Participant 12 who let his girlfriend drive the car. He had been instructed to turn off the engine for at least one minute in those cases when switching drivers in order to generate a new file for the new driver, but in practice this did not always occur. For the most part, the files created by the girlfriend could be found and discarded automatically via the eye tracking quality, because the face profile had not been created for her. Thus, the average eye tracking quality for her trips was much lower than for the intended participant. For those files where the driver was switched without shutting off the engine at all or not long enough, however, both drivers were logged in one file. This meant that the part logged for the girlfriend had to be removed manually, which was time consuming. The problem was aggravated by the fact that the participant's clothing in many cases disturbed the tracking, as mentioned above. One lesson that can be learned from this is that it cannot be expected that drivers will follow instructions that are given to them in a setting like the present one. It can be speculated that the drivers feel so familiar with the car after a while that they forget about instructions that do not immediately make sense or could be inconvenient, like turning off the engine for at least one minute. It is not necessarily the case that the drivers want to sabotage the study, rather on the contrary, the drivers are in most cases very helpful and return the car washed and cleaned, for example.

For Participant 14 the percentage of eye tracking and the total percentage of acceptable tracking were somewhat lower than for the other participants. A probable cause for that is that the participant scratched his nose a lot, thereby obstructing the view of his face.

Both clothing and behavioural habits that lead to a frequent obstruction of a view of the face are not very easy to control in a setting like the one used in this study. Except for trying to improve the logging equipment, not too much can be done about data loss resulting from circumstances like those. The participants had been asked not to wear sunglasses, because this was known to disturb the data logging. For the same reason only participants without beard and heavy mascara were selected. It appears very difficult, however, to ask the participants not to wear high collars or caps and the like when driving, and not to touch one's face, especially during a field study of such a duration, which is meant to capture naturalistic driving. A repositioning of the cameras or the addition of further cameras would most likely not have brought noticeable improvements.

The eye tracker tends to lose tracking in those occasions that are interesting from a distraction research point of view. When the drivers look away from the forward view and therefore from the cameras, eye tracking degrades. When the drivers cover their face with for example food, drinks, telephones and so on, tracking degrades as well. For research purposes it would be helpful to install more cameras, and they should be mounted especially in those locations that are known as distractors. It would be suitable to have one camera in a relatively low position in the middle console, another close to the right rear view mirror, one possibly even further back to cover glances to the hind seat row and out of the rear window, and one more that can cover over-the-shoulder glances to the left hand side. This would clearly aid the data analysis around the distraction period.

Driver distraction was not the only focus of this project. Drowsiness was included as well. In this report no analyses of the drowsiness data are included, because the drivers received only very few drowsiness warnings and they reported that they did not experience those as correct. In the light of these circumstances it did not appear meaningful to analyse the warnings any further, especially as the drowsiness detection algorithm itself was not disclosed to the analysis team. It is not likely that a driver who does not agree with a drowsiness warning given to him will take any action to rest.

11 General discussion

In this chapter more general aspects of the study, the AttenD algorithm and the distraction warnings will be discussed. Finally, recommendations for further analyses of the present data and suggestions for further research are presented.

11.1 Method and data treatment

Generally the method used in the present study appears to be suited for investigating driver distraction. Several thousand distraction cases were logged and more than 1,800 distraction warnings were issued during the treatment phases. The automatic eye tracking worked well over a prolonged time period without intervention by the experimenter. The quality of the tracking was constant over time, and the very varying lighting conditions did not adversely affect tracking. The data acquisition system worked as planned in principle, even though stability issues have to be solved in future studies, and a better power management is required for further and more large-scale studies.

In the present study the planned remote monitoring system could never be put into working order, due to software conflicts, which in the end led to the decision to drop live monitoring during the trials. Especially for the last two participants it would have been useful to have the monitoring active, because then the car could have been called in earlier for service and troubleshooting. For further studies it is highly recommended to have monitoring on a day-to-day basis, because early problem detection will not only lead to a reduction in data and time loss, but will also simplify analyses.

Generally, the experiences with the participants were positive, and the procedure used seemed to have worked for all involved. The participants generally felt comfortable with the setup and appreciated that the meetings were scheduled in a flexible manner, causing as little effort on the participants' side as possible. It was not assumed to influence the results whether the log system was activated when the participant visited the test centre VTI, or whether an experimenter met the participant at another location. It was deemed to be more important to ensure a smooth schedule for the participants to keep them cooperative. No participant complained about being filmed or asked about deleting any data. Some participants were quite interested in the research question, and the purpose of the study was explained to them in detail when they returned the vehicle. Other participants did not show as much interest in the study, but still cooperated willingly.

For this study the data analysis was done directly on the text files logged by the data acquisition system. There had been plans to build a relational database, but due to a number of unforeseen circumstances this had not been possible. Even though it worked well for the present amount of data to run the analysis in Matlab, it is recommended to build a database for quality assurance, traceability of results, backup facilities and for easy expansion. Some of the analyses performed here had a computational time of up to six hours when processing data from all participants. Increasing the number of participants would quickly lead to analysis durations that are not feasible. A database solution could be of help for larger datasets to reduce analysis time. Converting the files to binary format would also reduce the computation time since it is notoriously time consuming to read text files into memory.

The large variation between participants for many of the analysed performance indicators shows clearly that the number of participants was too low to draw any definite conclusions about the effect of the distraction warning system. This was aggravated by the fact that the data acquisition system became instable and a high data loss occurred for the last two participants. For further studies it is strongly recommended to run a larger number of participants, and to run them in parallel. The duration of the treatment phase with activated warning system, however, seemed to be appropriate. On the relatively rough analysis level used here no learning or adaptation curves that had not yet reached asymptote could be found. A more fine-grained analysis is necessary, though, before definite statements on a suitable duration for the baseline and the treatment phase can be made.

All analyses presented here were based either on single glance cases, meaning one data point at a time, or on the output of the AttenD algorithm employed. Fixations and saccades were not computed and analysed for the present data. The algorithm output was used, because one goal of the study was to investigate the performance of the algorithm. For glance duration it was of interest to know for how long in a row the driver glanced away from the FRD, regardless of whether this glance was the result of one single or several fixations. This could also have been computed using fixations, and it is recommended to work fixation based in the future, especially in order to validate the results obtained by using the algorithm output. The glance durations found with the analysis method used here correspond well to those found in the literature, which can be seen as an indication that AttenD works well in capturing glances off the forward roadway.

11.2 General results

The data from the present study suggest that the warnings do not have an immediate effect such that the driver looks up faster after having received a warning than without warning. It does not appear, either, that drivers avoid the warnings, because for most of the drivers the warning frequency did not decrease measurably. There is no increased concentration of glances in the field relevant for driving, or an increase in PRC values in the treatment condition. Neither does the standard distribution of gazes change from the baseline to the treatment phase. The only possible change could be found for the duration of individual glances, where the average percentage of very long glances decreased in the treatment phase. Furthermore, the number of times the attention buffer returned to zero before filling up completely after a warning decreased.

Thus, the results do not show large effects, but those that appear tend to go into the direction that a distraction warning has positive effects on driving safety and does not seem to have negative effects. The few effects that appeared did not change much from one treatment phase to the next. This might indicate that there is no noteworthy learning curve, even though it is too early for a definite statement on this question. The absence of a learning effect implies that the distraction warning has approximately the same effect already during the first experiences with the warning as it has later on.

The concern of system abuse exists especially for drowsiness warning systems – no warning might lead the driver to assume that there is no danger, instead of trusting one's one senses. The drivers might not stop and rest when they feel drowsy as they would have otherwise, because they either wait for a warning to tell them so, or worse, they expect the system to wake them up in time if they should fall asleep. Drivers might feel more self-confident and drive more often at night or when they do not feel fully alert. It is likely that this type of system misuse develops over time, when the driver gets to know his system and feels sure about what he can expect the system to do. A distraction warning system is not completely comparable to a drowsiness warning systems, even though the drowsiness warning system is designed to judge the driver state, just like the distraction warning system. Even though drowsiness varies over time, too, distraction usually has a much quicker onset and can disappear very quickly again. A distraction event can occur and disappear without the driver's noticing its existence, as long as nothing critical happens in the environment during the time of inattention.

The present data do not indicate any obvious misuse of the system, and theoretical considerations indicate that system misuse for this type of function is unlikely. It does not appear likely that a driver would consciously and calmly look away from the FRD until the warning comes, and by this exhaust the time until the warning comes. This is corroborated by the comments of several drivers that it felt like one needed to look away for a long time when trying to provoke a warning. Usually drivers did not feel comfortable when doing so. They were not explicitly asked after the study whether looking away felt more convenient with the system, but nobody mentioned anything possibly indicating such a phenomenon. It was not investigated whether drivers increased their speed with the distraction warning system activated, because there were too many uncontrolled variables to allow a meaningful analysis of speed change. It could be seen, though, that the percentage of glances of about two seconds duration did not increase, indicating that the participants did not start "waiting out" the warnings.

It appears worthwhile, to analyse further the frequency of distraction warnings in relation to different road types, and to consider the possibility to vary the sensitivity of the system in relation to road type. Even though the frequency of distraction warnings was found to be similar on all road types, it can be discussed whether distractions are equally dangerous on different road types. On motorways the environment is often rather simple and predictable, and lanes are wide and with forgiving roadsides. Speeds are high, however, which means that larger distances are covered during a given time, and crashes at high speeds can have severe consequences. On rural roads and especially in urban areas the environment is much more complex, constellations change faster, and the development of a traffic situation is much more unpredictable than on the motorway. However, speeds are lower, shorter distances are covered during a given time, and crashes at lower speeds are often less severe. It is much more likely, however, that unprotected road users are involved in crashes on urban and minor rural roads, and for them low speeds can already be fatal. This argumentation shows that it is not immediately clear whether a distraction warning system could afford to be more lenient at low or at high speeds.

11.3 Driver distraction warning system

Even though for very many crashes driver distraction is listed as at least a contributing factor (see e.g. Klauer, Dingus, Neale, Sudweeks & Ramsey, 2006; Wang, Knipling & Goodman, 1996), it can be discussed whether a distraction warning system in itself is a meaningful function. Is it necessary to warn a driver for looking away too much when there is in fact no danger, or is it enough to warn when there is imminent danger? It could be argued that a distraction warning system might teach the driver to look away less frequently, which would probably lead to a general increase in traffic safety. The present data, limited as they are, do not show a learning effect from the baseline phase to the treatment phase, and not either from the first week in the treatment phase to the last when it comes to warning frequency or reaction time. This does not imply that it is sure that there is no such effect – it could be hidden by the too small number of participants, it could show first after a longer learning and adaptation phase with the system, and it could be more subtle and only show in certain situations. It also appears likely from the present data that the 95th percentile in glance duration and the percentage of glances above 2 s decrease with the warnings active. This could be interpreted as an effect of the warnings on the total glance distribution and might be seen as a general improvement of glance behaviour. Such effects on general glance behaviour appear to be positive, and therefore it could be meaningful to warn for distraction for pedagogical reasons even if no immediate danger can be detected.

A further argument for warning under those circumstances, that is, with no imminent or latent danger present, would be that it can never be guaranteed that the sensors of the car can register all possible dangers, and by issuing a warning only when the vehicle detects another object close by, or a run-off-road, it is possible that a dangerous situation might be missed completely and therefore not be warned for.

On the other hand, too many warnings might quickly lead to irritated drivers who either want to switch off the system, or who do not take the warnings seriously any more. This would, of course, be a very negative consequence which outweighs the positive effects discussed before. During the approximately three weeks that the drivers in the present study experienced the system only one driver had complaints about the frequency of the warnings. If the AttenD algorithm were redesigned to reduce the warning frequency there should be hope that a large part of the driving population might accept the number of warnings.

Another possibility would be to connect the distraction warning system to other driver support systems. On the one hand, this sensor fusion would allow each system to take in more information in order to improve the functioning of each system, and it would, on the other hand, allow an adaptation of the warnings depending on the state of each system. Concretely this means that a combination of the distraction warning system, a curve speed/lane departure warning system and a forward collision warning system could work in the following way: The distraction warning system could take in information from the curve speed warning system on the curvature of the upcoming road stretch. If the road is judged to be very curvy, the threshold for the distraction warning system could be lowered. If the forward collision warning system then detects that an object is on collision course, and at the same time the driver's glance is not directed at the forward roadway, either a distraction warning could be issued immediately, or the forward collision warning could be issued earlier than in occasions where the driver is attentive. This communication between systems should decrease the occurrence of false alarms, and it should enhance acceptance if the warning strategy is well done and meaningful. This way it would be possible to extend the time for distraction warnings with no immediately sensed danger, and thus hopefully increasing acceptance for this type of warning. At the same time, in situations that are judged to be potentially dangerous the warning could come earlier, and in clearly dangerous situations the appropriate warning, for example forward collision, could be given while suppressing an additional inattention warning to avoid confusion in a critical situation. This would need to be tested both in a controlled environment and under realistic conditions.

The participants' comments showed that at least some of them had built quite complicated mental models of the AttenD algorithm, and in many cases their interpretations did not correspond to reality, but the mental models worked quite well anyway. Several participants realised that they tended to get warnings when they talked on the mobile phone, talked with a passenger or ate something while driving. They attributed the warning to their activity, which was not driving related. The eye tracker can, however, not detect such behaviour, and therefore it is not the eating or talking in itself that led to the warning, but the way the tracking works in such cases. Strong facial movements, which are typical for eating and talking, combined with a possible partial covering of the face often lead to degraded tracking performance or to misinterpretations of the tracking system. In cases in which eye tracking no longer is possible the algorithm reverts to head tracking, which is less precise and might lead to more “false” alarms. If the buffer already is depleted to 0.4 s, a loss in tracking will lead to a warning after those 0.4 s have passed. These factors might lead to warnings even though the driver looks straight ahead, which is not planned by the designers of the algorithm. These warnings are meaningful anyway, and interpreted as such by the participants in the study. For a more thorough discussion of the potential effects of different types of alarms see Lees and Lee (2007).

11.4 Recommendation for further analyses

The data analyses in this report are conducted on a relatively general level. It was deemed important to touch upon several aspects of behavioural changes, but there is a lot more information in the data, and it is important to go further with the analyses.

It has to be considered whether it is most meaningful to report mean values of for example glance duration, or whether other values should be considered. For driving behaviour that is deemed to be dangerous it can be of value to consider the extreme cases, like the longest glances, as they might be the most hazardous ones. In the present report this was done in Chapter 5 on glance duration, but extreme values of other variables are also of interest. With focusing on mean values the risk is run to hide extreme situations.

Another important aspect is to perform analyses on a more detailed level. The data could be filtered according to road type, time of day, time on task, weather, traffic density, passenger presence, use of nomadic and embedded devices, and so on, but all of those filtering would require a substantial amount of work. For road type the GPS log would need to be connected to the Swedish national road database, which is fully possible and would yield a vast amount of supplemental information on the current road. When coupling the GPS data to the road database data, however, it has to be made sure that the correct information from the day the trip was taken is used. Further enquiries have to be made to estimate the time effort required, how much of the information could be coded automatically, and whether manual coding is necessary. It is normal, for example, that it takes a few minutes for the GPS receiver to locate its current position, such that the first position logs would have to be removed before coupling the log to the database. In the present analyses speed was used as a substitute for road type, but it is only a rough approximation which could be greatly improved by having access to the actual road type.

Filtering for time of day does not pose a problem, but if time of day should reflect the lighting conditions, a matching has to be done with the changing times of sunrise and sunset over the year. For trip duration a simple approximation would be to use the time stamp that starts counting when a file is started, but for more accurate data it would be necessary to define how long an interruption between two trips has to be in order for a trip to be counted as separate from the preceding one.

Weather is not easy to log well. Either, the information might be gleaned from external sources that keep weather records, but this procedure is cumbersome, time consuming and not necessarily very accurate when it comes to determining the weather in a particular location on a road. Otherwise, the sensors on the vehicle can be used. Weather could be coded manually by watching the video, but this is not feasible both due to the amount of time required, and due to the fact that it is not always possible to discern the weather conditions well enough from the video. To some extent the information on the CAN bus could be used; windscreen wiper activity might indicate precipitation, the outside temperature can for example give information on rain versus snow, fog light activation might indicate fog, and so on. This would, however, only provide a rather rough and not always reliable classification.

Traffic density is difficult to derive from CAN measures. If no external information exists that can be coupled to a certain trip, it appears necessary to do the coding manually while watching the video. This is very cumbersome, as mentioned before. For analyses of this type it is recommended to use a dedicated software that allows real time or faster classification for variables like traffic density and weather.

The CAN data from the car used for the present study allowed to log whether something heavy was located on the passenger seat. This would in most cases be a person, but could be a heavy bag or the like. The presence of passengers in the back cannot reliably be determined by the data logged in this study. Passengers can influence the driver’s behaviour, however, which should not be completely neglected especially in research on driver distraction and drowsiness.

The operation of embedded devices might be determined via CAN signals, while video analysis is the only possible solution for most nomadic devices. For telephone use an radio frequency sensor might be of help. Even though it had been planned to use such a sensor in the present study, the idea was dropped due to problems with the technical installation. Even if such a sensor is present it is still necessary to determine whether the phone was used by the driver, a passenger or perhaps even somebody in a vehicle next to the target vehicle.

The variables listed above are only examples of filters that might be used to break down the data. All the variables mentioned might have an influence on driver distraction. Glance behaviour on the motorway might differ from glance behaviour on a rural road, a driver might get distracted more easily during the day than during the night or the other way around, and after having driven for a long time it might be more likely for a driver to become distracted. Inclement weather conditions might reduce visual distraction, the same might be true for dense traffic. Passengers in the car, on the other hand, might increase driver distraction, and conversations on the phone or the usage of other infotainment devices might have a similar effect.

It is highly recommended to do fixation based analyses on the eye tracking data. This way the validity of the AttenD algorithm can be analysed better, and it is also a common and valid approach to treat eye movement data (Salvucci & Goldberg, 2000). As mentioned, it was not done here, because the algorithm works gaze case based, and it was decided to use the same “raw” data format to evaluate the warning system. Furthermore, it would be necessary to decide which fixation computation algorithm to use, which in turn might influence the results.

Further glance based performance indicators could be employed, for example those developed in the SAVE-IT project (Zhang & Smith, 2004), which are partly based on those suggested by Recarte and Nunes (2000). A further interesting indicator might be the time spent on a secondary task, counting from the first interaction with or glance at a device until the final interaction with or glance at the device. Here it is necessary to define which time span can pass before a new secondary task is assumed to begin.

A further recommended treatment of the data is to analyse different events. It is of interest to know whether drivers choose when to be distracted, and whether they select the most appropriate moments. On the motorway it might be of interest to investigate whether there is a difference in glance behaviour depending on whether there is a ramp close by or not. The likelihood for being distracted in a car following situation versus a free driving situation is also of interest. The width of a rural road might influence visual distraction. The latter two are of special interest when the distraction warning system should be combined with a forward collision warning system or a lane departure warning system.

In the present study driver drowsiness was logged according to the algorithm provided by Continental VDO. It is important to relate driver distraction to other driver state variables, such as drowsiness, but also intoxication, for example. For sensor fusion and warning integration it is significant to know more about possible co variations of different driver states.

Furthermore, it is recommended to analyse the driving data in relation to the driver distraction occasions. Basically, first a change in driving behaviour can lead to a change in traffic safety. Sudden correcting steering wheel movements that come on the heels of a detected distraction might indicate a certain loss of control during the distraction period. Sudden brake operation and acceleration and jerk values might also indicate loss of control. Unfortunately in the present study no radar signal and no lane position signal was logged in the passenger car, because those sensors were not present in the vehicle and too expensive to install.

First analyses of driving behaviour data have been conducted and will be available in a separate publication (Kircher, A., Ahlstrom, C. & Kircher, K. submitted), and further analyses are planned.

11.5 Ground truth on driver distraction

Generally it would be very useful for evaluation purposes if a ground truth for the classification of driver distraction existed. A ground truth would be an independent source, which would provide definite information about whether the driver was distracted or not. With the ground truth available, the existing algorithm could be compared to modifications with respect to sensitivity and specificity in detection driver distraction. How to obtain the ground truth practically is another discussion. It could be a judgement by the driver him- or herself, or by an independent observer. However, there are difficulties with judgements based on introspection especially in the case of distraction. If a driver was truly distracted, it is not sure that he will be able to realise that a warning might have been relevant and justified, because he might not recall that he in fact looked away from the forward roadway for a substantial period of time. This became quite apparent during the pilot studies, and it was also reported by some of the participants. If a driver tried to provoke a distraction warning, it felt subjectively like it took quite an effort to look away from the road long enough to receive a warning. When the warnings occurred unprovoked, however, it happened that one felt that one had hardly looked away at all. Close analyses of the video material during the pilot revealed, however, that also for those cases the driver had looked away long enough for a warning to be justified.

Having an independent observer judge whether driver distraction is present or not holds other difficulties. Distraction is fleeting, it can be there in one moment and disappear in the next. If the observations have to be made on-line, the observer needs to be very attentive at all times in order to catch the occasions when the driver is distracted. If the categorisation is to be made off-line, it is necessary to have at least one video film of the driver with sufficient resolution and recording frequency. It would be better to have two films with different perspectives – one of the driver's face and eyes, the other from over the shoulder, to be able to observe his actions. High quality video recordings are storage intensive, which becomes less and less of a problem with increasing computer power, but for long term studies it is still a concern. No matter how advanced the technical equipment, observer based distraction ratings are still subjective to some degree and therefore not beyond any doubt.

An approved automated distraction measurement technique could be used as ground truth, but to the knowledge of the authors there exists no such undisputed technique yet. Such a technique would need to draw on physiological measures, which are often difficult to obtain. This is especially true for FOTs, where it is of paramount importance not to interfere with the drivers' daily routines, and where it is impossible to apply electrodes or the like on a participant.

11.6 Improvement of the AttenD algorithm

The distraction detection algorithm AttenD was used in its first version in the present study. In general the participants found the algorithm to be quite accurate and said that they could understand and accept many of the warnings. For a system acceptance it is very important, however, to reduce the amount of experienced false alarms as much as possible. At the same time the sensitivity, that is, the ability to detect true positives, should not be reduced. Ideally, the "experienced false alarms" should correspond as much as possible to what the system designer classifies as false alarms. In the present case, however, a warning while the eyes are directed at the road, but the driver talks on the telephone would be considered a false alarm by the system designer, while most of the participants experienced the warning as appropriate. The differentiation does not stop at experienced false alarms and experienced hits. A warning might very well be judged as coming after a distraction, but it might be a nuisance alarm when the driver already realised himself that he had been distracted and was just on his way to look back at the road. On the other hand, a warning in a situation like this might corroborate for the driver that the system actually works. Generally, nuisance alarms are considered as having less negative impact on acceptance as false alarms (Lees & Lee, 2007).

The participants received more warnings than it was intended when the algorithm was conceived. Even though the participants did not object or complain during the study, there is a great risk that acceptance levels will decrease over time with so many warnings. This implies that different ways to modify the algorithm in order to decrease the overall number of warnings and of course especially the false and nuisance warnings have to be found.

One possibility would be to extend the inhibition period after the last warning. It is possible that 15 s is not enough to complete a task that the driver is determined enough to complete to ignore a distraction warning. In-depth analyses of those video clips covering warnings that occurred with only 15–20 s in between could provide information as to whether the driver was still concerned with the same secondary task, or whether he had already initiated a new one. If those warnings that occur very soon after another warning were removed, the average inter-warning interval would increase substantially.

It is strongly suggested to put on an extra 15 s inhibition at the end of all CAN related inhibitions, meaning that when the car accelerates and crosses the 50 km/h limit, the first warning should not be able to come before at least another 15 s have passed. The same should be valid for steering and braking manoeuvres.

As mentioned, a substantial portion of the off-FRD glances was directed at those parts of the windscreen that were not part of the FRD. This might indicate that larger portions of the windscreen are, in fact, relevant for driving. It could, however, also indicate that drivers spend time looking at distracting objects outside of the vehicle. A thorough video based analysis would be necessary to determine which of the two is true. A suggestion to reduce the number of potential false alarms would be to include the complete windscreen in the FRD. On the other hand, this will most likely increase the probability of misses. For distraction it is very difficult to determine whether a warning would have been appropriate when no warning was given. Drivers are in many cases not aware of the fact that they have been distracted and can, thus, not give a meaningful introspective feedback, even if they wanted to. Only if the analysis is limited to cases in which other warning systems react, like for example a forward collision warning system or a lane departure warning system, it would be possible to determine whether distraction is a likely cause for this type of warning, and whether a distraction warning system might have prevented the warning. The same could be done for near crashes and other idiosyncrasies discovered in the driving data stream.

Another possibility to reduce the number of warnings would be to increase the radius of the FRD. This way fewer glances through the windows, but directed at targets far from the forward roadway would lead to a distraction warning. More advanced filters that take into account how the steering wheel is turned, how the road geometry looks like, and so on, might further improve the quality of the AttenD algorithm.

Increasing the buffer to a higher value than 2 s could also be an option. Furthermore, the increment and decrement of the buffer could be adapted depending on various variables related to CAN data or the environment, for example. More complex changes in the algorithm, like those mentioned in the end, would necessitate further studies to investigate which effect they would have, whereas the current data could be used to some extent in order to evaluate effects of simple inhibition changes.

So far the algorithm only considers visual distraction, that is, when the driver directs his gaze away from the forward roadway. It is well known, however, that cognitive distraction is a dimension that should not be neglected (Kircher, K., 2007). During cognitive distraction the driver’s eyes may very well be on the road, while the mind is not. There have been attempts to assess cognitive distraction via eye gaze behaviour (Victor, 2005), but this approach is not incorporated in the current AttenD algorithm. For future versions it is highly recommended, however, not to neglect this aspect of driver distraction, as the goal is to keep the mind on the traffic, and not only the eyes.

11.7 Future research

The most important recommendation for future research is to use a larger number of participants in order to be able to deal with the substantial between-subjects variation. It is difficult to state exactly how many participants should be used, because many variables influence this decision. It is likely, however, that more is known about which effect sizes can be expected for different advanced driver assistance systems, and how often certain types of warnings occur, when the SeMiFOT project (Victor & Gellerman, 2008) and the euroFOT project (euroFOT Consortium, 2008) are finished. These are two medium to large scale FOTs, which are under way at the time of writing. These projects employ continuous logging of a large number of variables, which will also provide valuable insights into normal driving patterns.

It could be interesting to require immediate feedback with respect to the experienced meaningfulness of a warning. This would most likely lead to an interference with naturalistic driving, but it could lead to valuable insights how the warnings are experienced by the driver. The feedback prompt might be given about ten or fifteen seconds or so after the last warning, in order to give the participant some time without losing the contiguity.

Requiring immediate feedback gives information about the “subjective truth”, as experienced by the driver. It would, however, also be very helpful to establish a ground truth for driver distraction. This is very difficult and a true challenge for further studies. It is thinkable that one needs to start in more controlled environments than real traffic with building up a solid benchmark for a distracted driver.

It is recommended to test different variations of the algorithm with respect to their effects on driving behaviour, driver behaviour, traffic safety and acceptance, in order to improve the detection of distraction occurrences and the rejection of false alarms.

A further topic of interest is to study the mental models the drivers create of the warning algorithm and investigate how they match with reality. A good match implies that drivers are not often surprised by unexpected warnings, which, in turn, is a good pre-condition for acceptance.

12 Conclusions

To summarise, it can be said that the distraction warning system employed in the present study was not found to influence driver behaviour on a general level. The drivers did not change their gaze behaviour much when the warnings were activated. They reported, however, that the warning system made them more aware of what they did while driving and one driver said that he actually changed his behaviour on the strategic level, avoiding telephone conversations and food consumption while driving. Therefore, it is too early to discount the present warning system as ineffective. More detailed analyses of eye gaze data and analyses of driving data are recommended.

The present method is considered to be a promising approach for driver distraction research, and the state-of-the-art eye trackers hold up to the expectations and demands. It has to be made sure, though, that the data acquisition system is robust enough to survive an extended period of time in a research vehicle.

References

Almén, L. (2003). *Reducing the effects of driver distraction: A comparison of distraction alerts on driver attention.* Linköpings universitet, Linköping.

Donmez, B., Ng Boyle, L. & Lee, J. D. (2006). The impact of distraction mitigation strategies on driving performance. *Human Factors, 48*(4), 785–840.

Donmez, B., Ng Boyle, L. & Lee, J. D. (2007). Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis & Prevention, 39*(3), 581–590.

Donmez, B., Ng Boyle, L., Lee, J. D. & McGehee, D. V. (2006). Drivers' attitudes toward imperfect distraction mitigation strategies. *Transportation Research Part F, 9*, 387–398.

euroFOT Consortium. (2008). euroFOT - Bringing intelligent vehicles to the road. http://www.eurofot-ip.eu/.

Karlsson, R. (2005). *Evaluating driver distraction countermeasures* (VTI notat No. 28A-2005). Linköping: VTI.

Kircher, A., Ahlstrom, C. & Kircher, K. (submitted). *Gaze-based driver distraction warning system - relationship between gaze data and steering wheel data.* Paper presented at the 16th ITS World Congress 2009, Stockholm, Sweden.

Kircher, K. (2007). *Driver distraction: A review of the literature* (VTI rapport No. 594A). Linköping, Sweden: VTI (Swedish National Road and Transport Research Institute).

Kircher, K., Kircher, A. & Claezon, F. (2009). *Distraction and drowsiness. A field study* (VTI rapport No. 638A). Linköping: VTI.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. & Ramsey, D. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (Technical Report No. DOT HS 810 594). Washington DC: NHTSA.

LeBlanc, D., Sayer, J. R., Winkler, C., et al. (2006). *Road departure crash warning system field operational test: Methodology and results. Volume I: Technical Report* (Technical Report No. UMTRI-2006-9-1). Ann Arbor, MI, USA: The University of Michigan Transportation Research Institute (UMTRI).

Lees, M. N. & Lee, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics, 50*(8), 1264–1286.

Recarte, M. A. & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied, 6*(1), 31–43.

Salvucci, D. D. & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye tracking protocols.* Paper presented at the Eye Tracking Research and Applications Symposion, Palm Beach Gardens, FL, USA.

SmartEye AB. (2007). User Manual; Revision 199. Gothenburg, Sweden.

Tijerina, L., Johnston, S., Parmer, E. B., Winterbottom, M. D., & Goodman, M. J. (2000). *Driver distraction with wireless telecommunications and route guidance systems* (Technical Report No. DOT HS 809-069). Washington DC: US Department of Transportation, National Highway Traffic Safety Administration.

Tsimhoni, O. (2003). *Time-sharing of a visual in-vehicle task while driving: findings from the task occlusion method* (Technical Report No. UMTRI-2003-13). Ann Arbor, MI: The University of Michigan Transportation Research Institute (UMTRI).

Victor, T. (2005). *Keeping eye and mind on the road.* Unpublished Dissertation, University of Uppsala, Uppsala.

Victor, T., & Gellerman, H. (2008). SeMiFOT - A SAFER project. Sweden Michigan Field Operational Test. Information Brochure. https://www.chalmers.se/safer/EN/projects/traffic-safety-analysis/semifot/downloadFile/attachedFile_f0/Folder_SeMiFOT_project?nocache=1201267213.94.

Victor, T., Harbluk, J. L. & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F, 8*(2), 167–190.

Wang, J.-S., Knipling, R. R. & Goodman, M. J. (1996, 7–9 October 1996). *The role of driver inattention in crashes: new statistics from the 1995 crashworthiness data system.* Paper presented at the 40th Annual Proceedings Association for the Advancement of Automotive Medicine, Vancouver, British Columbia.

Zhang, H. & Smith, M. R. H. (2004). *A final report of safety vehicles using adaptive interface technology (Phase I: Task 7: Visual Distraction Research)* (Final Report): Delphi Electronics & Safety.

www.vti.se
vti@vti.se

VTI är ett oberoende och internationellt framstående forskningsinstitut som arbetar med forskning och utveckling inom transportsektorn. Vi arbetar med samtliga trafikslag och kärnkompetensen finns inom områdena säkerhet, ekonomi, miljö, trafik- och transportanalys, beteende och samspel mellan människa-fordon-transportsystem samt inom vägkonstruktion, drift och underhåll. VTI är världsledande inom ett flertal områden, till exempel simulatorteknik. VTI har tjänster som sträcker sig från förstudier, oberoende kvalificerade utredningar och expertutlåtanden till projektledning samt forskning och utveckling. Vår tekniska utrustning består bland annat av körsimulatorer för väg- och järnvägstrafik, väglaboratorium, däckprovnings-anläggning, krockbanor och mycket mer. Vi kan även erbjuda ett brett utbud av kurser och seminarier inom transportområdet.

VTI is an independent, internationally outstanding research institute which is engaged on research and development in the transport sector. Our work covers all modes, and our core competence is in the fields of safety, economy, environment, traffic and transport analysis, behaviour and the man-vehicle-transport system interaction, and in road design, operation and maintenance. VTI is a world leader in several areas, for instance in simulator technology. VTI provides services ranging from preliminary studies, highlevel independent investigations and expert statements to project management, research and development. Our technical equipment includes driving simulators for road and rail traffic, a road laboratory, a tyre testing facility, crash tracks and a lot more. We can also offer a broad selection of courses and seminars in the field of transport.

vti

HUVUDKONTOR/HEAD OFFICE
LINKÖPING
POST/MAIL SE-581 95 LINKÖPING
TEL +46(0)13 20 40 00
www.vti.se

BORLÄNGE
POST/MAIL BOX 760
SE-781 27 BORLÄNGE
TEL +46 (0)243 446 860

STOCKHOLM
POST/MAIL BOX 55685
SE-102 15 STOCKHOLM
TEL +46 (0)8 555 770 20

GÖTEBORG
POST/MAIL BOX 8077
SE-402 78 GÖTEBORG
TEL +46 (0)31 750 26 00