

# Investigation of driver sleepiness in FOT data

Final report of the project SleepEYE II, part 2



## Authors

Carina Fors, VTI

David Hallvig, VTI

Emanuel Hasselberg, Smart Eye

Jordanka Kovaceva, Volvo Cars

Per Sörner, Smart Eye

Martin Krantz, Smart Eye

John-Fredrik Grönvall, Volvo Cars

Anna Anund, VTI



Virtual Prototyping and  
Assessment by Simulation

[www.vipsimulation.se](http://www.vipsimulation.se)



## Preface

*SleepEYE* was a collaborative project between Smart Eye, Volvo Cars and VTI (the Swedish National Road and Transport Research Institute) within the competence centre *Virtual Prototyping and Assessment by Simulation* (ViP). The project was carried out during the years 2009–2011, and included development of a camera-based system for driver impairment detection and development of a driver sleepiness classifier adapted for driving simulators.

A continuation project called *SleepEYE II* was initiated in 2011. This project has included three work packages: 1) simulator validation with respect to driver sleepiness, 2) assessment of driver sleepiness in the euroFOT database using the camera system that was developed in *SleepEYE*, and 3) further development and refinement of the camera system. The present report documents the work that has been undertaken in work packages 2 and 3 of *SleepEYE II*.

John-Fredrik Grönvall at Volvo Cars has been the project manager, and Jordanka Kovaceva, also Volvo Cars, has been responsible for all tasks related to the euroFOT database.

Martin Krantz, Emanuel Hasselberg and Per Sörner at Smart Eye have done all work in work package 3 and also contributed to the data processing in work package 2.

Anna Anund, Carina Fors and David Hallvig at VTI have been responsible for developing and evaluating the methodologies used in work package 2. Beatrice Söderström, VTI, recruited participants to the video experiment.

Katja Kircher at VTI, Julia Werneke and Jonas Bårgman at SAFER (the Vehicle and Traffic Safety Centre at Chalmers) have reviewed the report and given valuable feedback on the content.

The *SleepEYE II* project was financed by the competence centres ViP and SAFER (Vehicle and Traffic Safety Centre at Chalmers).

Thanks to all of you who have contributed to the project.

Linköping, December 2012

*Anna Anund*

## Quality review

Peer review was performed on 26 October 2012 by Julia Werneke and Jonas Bårgman, SAFER (the Vehicle and Traffic Safety Centre at Chalmers) and on 9 November 2012 by Katja Kircher, VTI. Carina Fors has made alterations to the final manuscript of the report. The ViP Director Lena Nilsson examined and approved the report for publication on 27 June 2013.

## Table of contents

|  |    |
|--|----|
| Summary .....  | 5  |
| 1 Introduction .....   | 7  |
| 1.1 The SleepEYE I project.....                                    | 9  |
| 1.2 The euroFOT database.....                                      | 10 |
| 1.3 Aim.....   | 10 |
| 1.4 Report structure .....   | 10 |
| 1.5 Related publications .....                                     | 11 |
| 2 Improvements of the embedded camera system .....                 | 13 |
| 2.1 Extended head rotation range.....                              | 13 |
| 2.2 Improved initialization logic .....                            | 13 |
| 2.3 Improved tracking recovery.....                                | 13 |
| 3 Evaluation of blink data quality .....                           | 15 |
| 3.1 Evaluation of blink data from the camera system .....          | 15 |
| 3.2 New blink detection algorithm .....                            | 21 |
| 3.3 Post processed blink data.....                                 | 25 |
| 3.4 Summary and conclusions.....                                   | 32 |
| 4 Sleepiness indicators: selection, performance and analysis ..... | 33 |
| 4.1 SleepEYE I field data .....                                    | 33 |
| 4.2 From SleepEYE field data to euroFOT data .....                 | 37 |
| 5 Video ORS.....   | 41 |
| 5.1 Method.....  | 41 |
| 5.2 Results.....   | 44 |
| 5.3 Discussion .....   | 47 |
| 6 Analysis of driver sleepiness in euroFOT data.....               | 49 |
| 6.1 Method.....  | 49 |
| 6.2 Results.....   | 49 |
| 6.3 Discussion .....   | 50 |
| 7 Summary and recommendations .....                                | 53 |
| 7.1 Lessons learned .....  | 54 |
| 7.2 Recommendations for future FOT studies .....                   | 54 |
| 7.3 Contribution to ViP .....                                      | 55 |
| References .....   | 56 |

## Appendix A – Video-ORS instructions



## Investigation of driver sleepiness in FOT data

by Carina Fors<sup>1</sup>, David Hallvig<sup>1</sup>, Emanuel Hasselberg<sup>2</sup>, Jordanka Kovaceva<sup>3</sup>, Per Sörner<sup>2</sup>, Martin Krantz<sup>2</sup>, John-Fredrik Grönvall<sup>3</sup>, Anna Anund<sup>1</sup>

<sup>1</sup>VTI (Swedish National Road and Transport Research Institute)

<sup>2</sup>Smart Eye

<sup>3</sup>Volvo Cars

### Executive summary

Driver sleepiness contributes to a great number of motor vehicle accidents every year. In order to reduce the number of sleepiness related accidents, more knowledge on e.g. prevalence, countermeasures and driver behaviour is needed. Data from field operational tests (FOT) has a potential to provide such knowledge with high ecological validity.

The objective of the project was to propose and evaluate methods for identification of driver sleepiness in FOT data. More specifically, the aim was to identify objective indicators of sleepiness – based on driving behaviour, eye blink behaviour and models of circadian rhythm – and to evaluate a subjective video scoring method for estimating driver sleepiness levels. Data from two separate projects were used: 1) the ViP-project SleepEYE, in which a controlled field test was conducted, and 2) euroFOT, which was a large scale FOT.

In a first step the data quality of blink-based indicators obtained from a camera system was evaluated. It was concluded that the data quality had to be improved and thus, a new detection algorithm was devised and implemented. The new detection algorithm had an acceptable detection rate (approximately 50 %) when applied to data from the SleepEYE field test, but for euroFOT data the number of identified blinks was very low (< 5 blinks/min) in about half of the trips. There is thus a need for further improvements of the blink detection algorithm.

An in-depth study on indicators of driver sleepiness was carried out using data collected in the SleepEYE experiment, with the purpose of employing the best indicators to study driver sleepiness in the euroFOT database. The most promising indicators were found to be mean blink duration and number of line crossings. A sleepiness classifier was suggested based on the distribution of the data (i.e. visual inspection). When applied to SleepEYE data the classifier was found to have good specificity while the sensitivity of the classifier was not so good. From euroFOT no true data on the drivers' sleepiness levels were available and it was therefore not possible to evaluate the performance of the classifier. However, an explorative analysis showed that only very few data points were classified as sleepy. This may be reasonable since most trips were conducted during daytime, but it is a somewhat disappointing result for the project.

A study was carried out on whether it is possible to use video recordings of drivers in order to estimate the drivers' self-rated level of sleepiness. Forty participants rated 54 one-minute video clips of an equal number of sleepy and alert drivers on a scale with three levels (alert, first signs of sleepiness, very sleepy). The results of the study showed that performing such observer rated sleepiness (ORS) estimations on drivers is extremely difficult. The videos available in FOTs are usually of rather poor quality

which, clearly limits the possibility of making reliable observer rated sleepiness estimations.

In conclusion, studying driver sleepiness in (existing) FOT data is difficult, for several reasons: 1) eye camera based indicators suffer from detection errors and low detection rate, 2) driving-based indicators are influenced by e.g. road curvature and traffic density, 3) models of sleepiness cannot be used since no information on hours slept and time awake is available, and 4) video scoring is not reliable, at least not given the quality of the available video recordings.

In future studies on driver sleepiness in FOTs sleepiness should be addressed in the FOT design. Some information about the drivers' sleep and sleepiness (ratings, sleep diaries, etc.) must be collected during the test; otherwise it will be very difficult to get any useful results.



# 1 Introduction

Driver sleepiness contributes to a great number of motor vehicle accidents (NTSB 1999; Åkerstedt 2000). In the UK the proportion of sleep related vehicle accidents has been estimated to about 10-20%, where the higher percentage refers to motorway (Horne & Reyner 1995; Maycock 1997). A study carried out in France found that, approximately, 10% of single vehicle accidents are related to sleepiness (Philip, Vervialle et al. 2001) whereas studies in Finland estimate that sleepiness is a contributing factor in 15% of fatal accidents caused by nonprofessional drivers (Radun & Summala 2004). A recent in-depth study of databases of crashes in the US between 1999 and 2008 concluded that an estimated 7% of all crashes and 16.5% of fatal crashes involved a sleepy driver (Tefft 2012). Compared to other police reported crashes, sleep related crashes are associated with a higher risk of death and severe injury (Horne & Reyner 1995).

A fundamental problem when studying driver sleepiness is, in fact, how to measure sleepiness. In short, the essence of the problem is that although an intuitive understanding of sleepiness is rather self-evident to most people, the scientific concept is not so easily defined and it is not possible to directly measure sleepiness (Cluydts, De Valck et al. 2002). Sleepiness is therefore regarded as a hypothetical construction, which leads to the problem of operationalization (Maccorquodale & Meehl 1948). Different approaches to measuring sleepiness have been proposed and considered in the literature on driver sleepiness. According to Liu et al. (2009), some of the most commonly studied approaches are: (1) subjective sleepiness estimates, (2) physiological measures, and (3) driving behaviour based measures.

Most research on driving while sleepy has been done in driving simulators (Liu, Hosking et al. 2009). In such driving simulator studies it has, for example, been found that subjective self-rated sleepiness is higher after sleep loss and during night-time driving (Gillberg, Kecklund et al. 1996; Åkerstedt, Peters et al. 2005) as well as that driver sleepiness is associated with (and likely causes) increased lateral variability (Anund, Kecklund et al. 2008; Lowden, Anund et al. 2009) which also results in an increased number of line crossings (Reyner & Horne 1998; Ingre, Åkerstedt et al. 2006). The absence of any *real* risk in the driving simulators is of course one of the major reasons why sleepiness is studied in simulators. However, the absence of any real risk may also be a drawback since it may have an influence on driving behaviour, and therefore make generalization, of results found in simulator studies to real world driving, invalid. This has motivated the studying of validity of driving simulators with regard to driver sleepiness. Although the results from various validation studies are somewhat different, sleepiness and the effects of sleepiness tend to be more pronounced in the simulator than on the real road, except for extremely long driving sessions, which might be explained by ceiling effects (Philip, Sagaspe et al. 2005b; Sandberg, Anund et al. 2011b; Davenne, Lericollais et al. 2012).

Driving performance and sleepiness have also been studied in controlled experiments on real roads, where sleep deprivation has been found to be associated with increased subjective sleepiness (Philip, Sagaspe et al. 2005a; Sandberg, Anund et al. 2011a) and an increased number of line crossings (Philip, Sagaspe et al. 2005a; Sagaspe, Taillard et al. 2008). Driver sleepiness has also been shown to have a relationship with electroencephalographic (EEG) activity (Kecklund & Åkerstedt 1993; Sandberg, Anund et al. 2011a; Simon, Schmidt et al. 2011), blink duration (Häkkinen, Summala et al. 1999; Sandberg, Anund et al. 2011a) and speed (Sandberg, Anund et al. 2011a). An

overview of indicators and methods for detection of driver sleepiness can be found in a previous SleepEYE report (Fors, Ahlström et al. 2011).

A limitation of the research methodologies mentioned above is that they either rely on drivers' self-reported experiences of sleepy driving or reflect driving behaviour in an artificial setting where the driver is being monitored. To what extent the results from such studies can be generalized to real world driving is not fully known. Therefore, other research methods are needed in order to validate and to supplement these results.

In recent years, several large scale naturalistic driving studies (NDS) and field operational tests (FOT) have been conducted to better understand drivers' behaviour. In these studies drivers are unobtrusively monitored in a natural driving setting, often for a relatively long time. The objectives of NDS/FOT studies are often to study the effects of driver assistance systems or to analyse accidents and incidents. However, NDS/FOT studies are also of interest from a sleepiness perspective, for several reasons. First, driver behaviour can be investigated without any influence from a test leader or a test protocol. Second, risk factors and prevalence of driver sleepiness could potentially be studied. It might also be possible to identify countermeasures used by sleepy drivers and to evaluate the effect of those.

Since NDS/FOT studies are not controlled experiments, conventional analysis methods cannot be used. A key issue is the lack of reference values (e.g. self-reported subjective sleepiness) or well-defined conditions (e.g. alert vs. sleep deprived). Thus, in order to use NDS/FOT data for sleepiness research, new and/or modified analysis methods are needed.

There have been a few attempts to investigate driver sleepiness using NDS/FOT data reported in the literature. Barr and colleagues have identified and characterized driver sleepiness among truck drivers in a naturalistic driving study (Barr, Yang et al. 2005). The entire video library was reviewed by a video analyst who rated the drivers' sleepiness levels on a five-point scale. A subset of the sleepiness events was then analysed in detail. A different methodology was used by Dingus and colleagues (Dingus, Neale et al. 2006). Instead of reviewing all video data a number of triggers were defined and implemented in the data acquisition system. Whenever a trigger criterion was fulfilled, the acquisition system recorded video and driving data for a short time period of 90 s before and 30 s after the triggering event. The collected videos were then reviewed by video analysts who categorized the events and rated the drivers' sleepiness levels. The rating scale used in both studies was based on the Observer Rating of Drowsiness (ORD) proposed by Wierwille and Ellsworth (Wierwille & Ellsworth 1994). The ORD scale is a 5-grade scale ranging from "Not drowsy" to "Extremely drowsy". There are two major limitations with the ORD scale (see also Chapter 5): 1) there is no guidance on how to use and interpret the scale, i.e. the ratings will depend on the rater's own interpretation, and 2) the scale has not been validated to any "true drowsiness level". There is thus a need for a validated video scoring method.

The objective of the present project, SleepEYE II, was to propose and evaluate methods for the analysis of driver sleepiness in FOT data. More specifically, the aim was to identify objective indicators of sleepiness and evaluate a subjective video scoring method for estimating driver sleepiness levels. Data from two separate projects were used: 1) SleepEYE, in which a controlled field test was conducted, and 2) euroFOT, which was a large scale field operational test conducted in the years 2008–2011.

The SleepEYE II project included three work packages. Work package 1 aimed at conducting a simulator validation study using data from the SleepEYE experiments.

The results from that study are presented in a separate report (Fors, Ahlström et al. 2013). Work package 2 concerned the methodologies for analysing driver sleepiness in FOT data, as described above. In work package 3, the embedded 1-camera system that was developed in SleepEYE was further improved. The present report constitutes the documentation of work packages 2 and 3.

## 1.1 The SleepEYE I project

The present project is a continuation of the SleepEYE project (which will be referred to as SleepEYE I from now on). The aims of SleepEYE I were to develop and evaluate a low cost eye tracker unit, to identify sleepiness indicators feasible for use in driving simulators, to determine indicator thresholds for sleepiness detection, and to combine the indicators into a simple classifier (Fors, Ahlström et al. 2011).

SleepEYE I began with a literature review on indicators for driver sleepiness and distraction. The following parts of the project were then focused on driver sleepiness. The most promising indicators for use in driving simulators were found to be blink duration, percentage of eyelid closure (PERCLOS), blink frequency, lateral position variation, and predicted sleepiness level based on a mathematical model of wakefulness and sleep. The eye related indicators were implemented in a 1-camera embedded eye tracker unit that was especially designed to be used in vehicles.

The project included two experiments. The first was a field test where 18 participants took part in one alert and one sleepy driving session (as defined by time of day) on a motorway. 16 of the 18 participants also participated in the second experiment which was a simulator study similar to the field test. Data from the eye tracker, from physiological sensors, and from the vehicle were continuously registered during the driving sessions.

The field test data were used for evaluation of the 1-camera system with respect to the sleepiness indicators. Blink parameters from the 1-camera system were compared to blink parameters obtained from a reference 3-camera system and from the electrooculogram (EOG). It was found that the 1-camera system missed many blinks and that the blink duration was not in agreement with the blink duration obtained from the EOG and from the reference 3-camera system. However, the results also indicated that it should be possible to improve the blink detection algorithm since the raw data looked well in many cases where the algorithm failed to identify blinks.

The sleepiness classifier was created using data from the simulator experiment. In a first step, variants of the indicators identified in the literature review were implemented and evaluated. The most promising indicators were then used as inputs to the classifier.

The final set of indicators was thus used to estimate a sleepiness level. The indicators were based on the Karolinska sleepiness score (KSS) value the driver reported before the driving session ( $KSS_{estSR}$ ), standard deviation of lateral position ( $SDLP$ ), and fraction of blinks  $> 0.15$  s ( $fracBlinks$ , for EOG-based and 1-camera based). An optimal threshold for discriminating between KSS above and below 7.5 was determined for each indicator. The performances were in the range of 0.68-0.76.

Two decision trees based on the selected indicators were created: one using  $fracBlinks_{EOG}$  and one using  $fracBlinks_{1CAM}$ . The performances of the two trees were 0.82 and 0.86, respectively (on the training dataset), i.e. the overall performance of the EOG based and the 1-camera based classifier were similar, although individual differences could be seen. The performance decreased to 0.66 when using a validation

dataset from another study, which illustrates the difficulties in creating a generalized sleepiness classifier.

A detailed description of SleepEYE I can be found in the project report (Fors, Ahlström et al. 2011).

## 1.2 The euroFOT database

The euroFOT project was the first large scale European field operational test on active safety systems. The project included 28 partners and test fleets in four different countries. Data have been collected from hundreds of vehicles during one year. The Swedish test fleet consisted of 100 passenger cars and 50 trucks, and it was operated by Volvo Cars, Volvo Technology and Chalmers. In the present project data from the Swedish passenger cars are used.

Data were collected during the private and occupational everyday driving by ordinary drivers. The data come from several sources, such as the in-vehicle data network (CAN), GPS, eye tracking, and video of the driver and vehicle surroundings (the context). The software of the eye tracking system was upgraded during the course of the test. Blink detection was added in the end of the project, why only data from the last upgrade – version 21 – are relevant for the present project. From a total of 66750 trips in the (Swedish passenger car) euroFOT database 2442 trips satisfied the software version condition. The data come from 36 different drivers and correspond to 6836 minutes of driving.

Since euroFOT was a naturalistic study no data on the drivers' sleepiness levels are available. The reason is mostly due to a wish to make the driving as natural as possible and to avoid interaction with the drivers. Using electrodes to measure sleepiness, or asking questions, or other types of obtrusive measures are therefore out of the question.

## 1.3 Aim

The aims of SleepEYE II, work packages 2 and 3, were to:

- Improve head rotation tracking and initialization logic for the camera system.
- Propose and evaluate sleepiness indicators suited for FOT data.
- Develop and evaluate a methodology for video scoring of driver sleepiness.

## 1.4 Report structure

The report begins with a chapter about the work that was done to improve the software of the camera system (Chapter 2). In Chapter 3.1, the blink data obtained from the new improved camera software is evaluated. It is concluded that the blink detection needs to be improved, which is done in Chapter 3.2. The blink data quality from the new algorithm is evaluated in Chapter 3.3. Blink parameters and some other sleepiness indicators are evaluated in Chapter 4. In Chapter 5 the video scoring method is described and evaluated. In Chapter 6 the developed and suggested methods are applied to euroFOT data (however, the video scoring method turned out to give very poor results and it was thus never used on euroFOT data). The methods are discussed, and some conclusions and recommendations for future studies are given in Chapter 7 shows

a flow chart of the work that has been undertaken within the project, and how the work has been structured in the report.

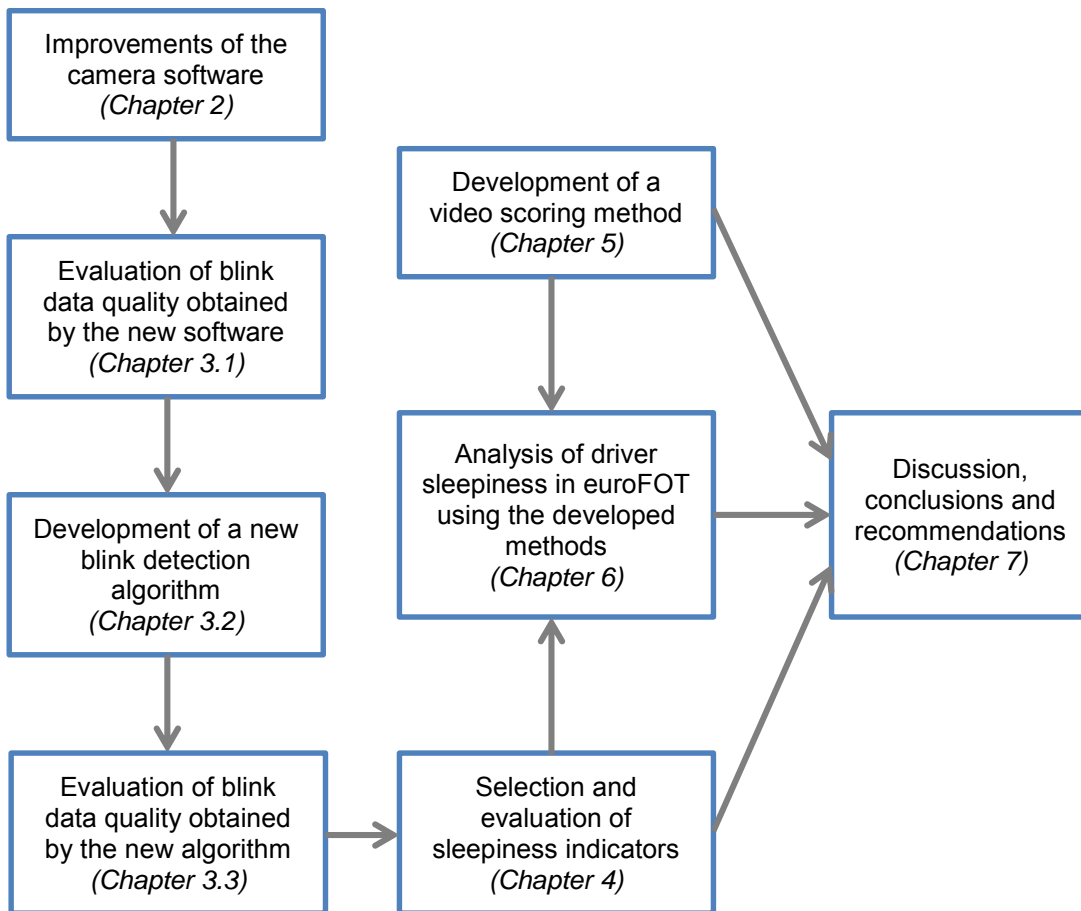


Figure 1: The flow chart shows the work that has been undertaken within the project, and how the work has been structured in the report.

## 1.5 Related publications

SleepEYE I resulted in three publications:

- Fors C, Ahlström C, Sörner P, Kovaceva J, Hasselberg E, Krantz M, Grönvall J-F, Kircher K, Anund A (2011). *Camera-based sleepiness detection – Final report of the project Sleep EYE*. ViP publication 2011-6, The Swedish National Road and Transport Research Institute, Linköping, Sweden.
- Ahlstrom C, Kircher K (2010). *Review of real-time visual driver distraction detection algorithms*. 7<sup>th</sup> International Conference on Measuring Behaviour, Eindhoven, Netherlands, 2010.
- Ahlstrom C, Kircher K, Sörner P (2011). *A field test of eye tracking systems with one and three cameras*. 2<sup>nd</sup> International Conference on Driver Distraction and Inattention, Gothenburg, Sweden, 2011.

The first publication is the project report describing the experiments, the results of the camera evaluation with regard to blink parameters, and the sleepiness detector for driving simulators. The two following publications were produced outside the project, but are based on data from the project. In the first of these two publications the 1-camera system is compared with the 3-camera system with respect to gaze parameters. The second publication is based on the literature review on distraction indicators that was done within the project.

In SleepEYE II the work done in Work Package 1 has been published in a separate report:

- Fors C, Ahlström C, Anund A (2013). *Simulator validation with respect to driver sleepiness and subjective experiences – Final report of the project SleepEYE II, part 1*. ViP publication 2013-1, The Swedish National Road and Transport Research Institute, Linköping, Sweden.

In addition, the results of SleepEYE II were presented at a workshop in Gothenburg in June 2012. About 15 people from SAFER, Smart Eye, Volvo Cars and VTI participated in the workshop.

## 2 Improvements of the embedded camera system

Eyelid estimation is not possible without head tracking, and even if there is head tracking good estimation of the position and rotation of the head is needed for high quality eyelid measurements.

The existing software in the camera system has problems estimating the depth of the head position (i.e. the distance from the camera to the head). This is due to the fact that it is a 1-camera system. The problem makes it hard to track the rotation of the head because the position predicted by the system doesn't correlate with the real position. Therefore it is hard to track horizontal head rotations larger than 10–15 degrees.

A solution would be to adapt the system during tracking, i.e. when there is more information. This could unfortunately not be done on the embedded system due to the limited computational power.

Instead a solution where more tracking points were added was chosen, which gives a more robust estimation of the head rotation even if the estimated depth position of the head is not perfect.

### 2.1 Extended head rotation range

In order to extend the head rotation range beyond frontal and near frontal head poses, more feature tracking points need to be added. Good feature tracking points are corner-like, i.e. they can be identified both vertically and horizontally in the camera image.

To find better tracking features for eyebrows and ears a new corner detector was implemented. Due to the large computational requirements of the detector, the majority of time went into optimization. The resulting detector was able to find good corner points in a local region.

A significant effort also went into improvements of the head pose estimation algorithm, especially regarding its sensitivity to outliers, for example partial occlusions and mouth movements.

Improvements in other areas related to tracking have also been done.

Taken together, these and other improvements increased the horizontal head rotation range by approximately 10–20 degrees, depending on subject.

### 2.2 Improved initialization logic

The initialization mode of the eye tracker was improved and resulted in a much shorter time to tracking, which in turn improved the ability to initialize and track on highly mobile subjects. Most of the work was done to the state machine and lowered the requirements for a stationary head, without losing any tracking quality.

### 2.3 Improved tracking recovery

Tracking recovery is the state the eye tracker goes into when head tracking is lost, usually because of major occlusions or large head rotations. Tracking recovery is fast since a model of the head already exists, and is therefore preferred over initialization.

For different reasons, the tracking recovery cannot always recover and a new initialization is required. The time delay before recovery switches over to initialization

was previously constant. Improvements have now been made so that the time adapts to the length of time the eye tracker previously had good tracking of the face, balancing quick trashing of failed initializations against the survival of good tracking models.

Improvements were also made to the tracking recovery algorithms, resulting in a quicker recovery.



### 3 Evaluation of blink data quality

The software of the embedded 1-camera system that was developed and evaluated with regard to blink parameters in SleepEYE I was further improved in the present project (see Chapter 2). The improved version was installed in the euroFOT cars during the last 1–2 months of the experiment. In order to have the best data possible, only data from that period were included in the present project.

As a first step, the performance of the new improved software version was evaluated with regard to its ability to detect blinks and to determine blink duration.

#### 3.1 Evaluation of blink data from the camera system

##### 3.1.1 SleepEYE field data

The 1-camera data that were acquired during the field test in SleepEYE I were re-run with the new version of the camera software and compared to blink data from the EOG. Relevant blink parameters were obtained from the EOG by the LAAS<sup>1</sup> algorithm (Jammes, Sharabaty et al. 2008). In SleepEYE I the results showed that the LAAS algorithm had a detection rate of about 98%, which was considered to be good enough as a "ground truth" for evaluation of the camera blink detection algorithm.

Figure 2 shows the number of blinks detected by LAAS and by the improved camera system per participant and driving session. On average, the camera system detects 39% of the number of blinks detected by LAAS (day: 41%, night: 36%). In two cases the camera system detects more blinks than LAAS. By comparing the camera detected blinks with the EOG, it is obvious that the camera data contains a lot of erroneous detections in these particular cases.

---

<sup>1</sup> LAAS is an abbreviation for *Laboratoire d'Analyse et d'Architecture des Systèmes (Toulouse, France)*, which is the research unit where the algorithm was developed.

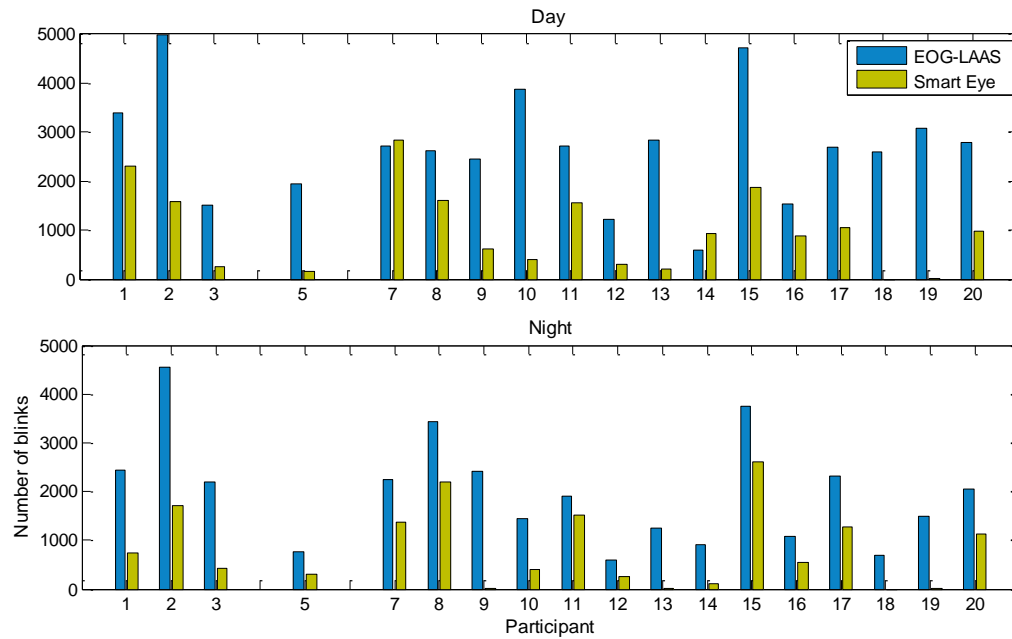
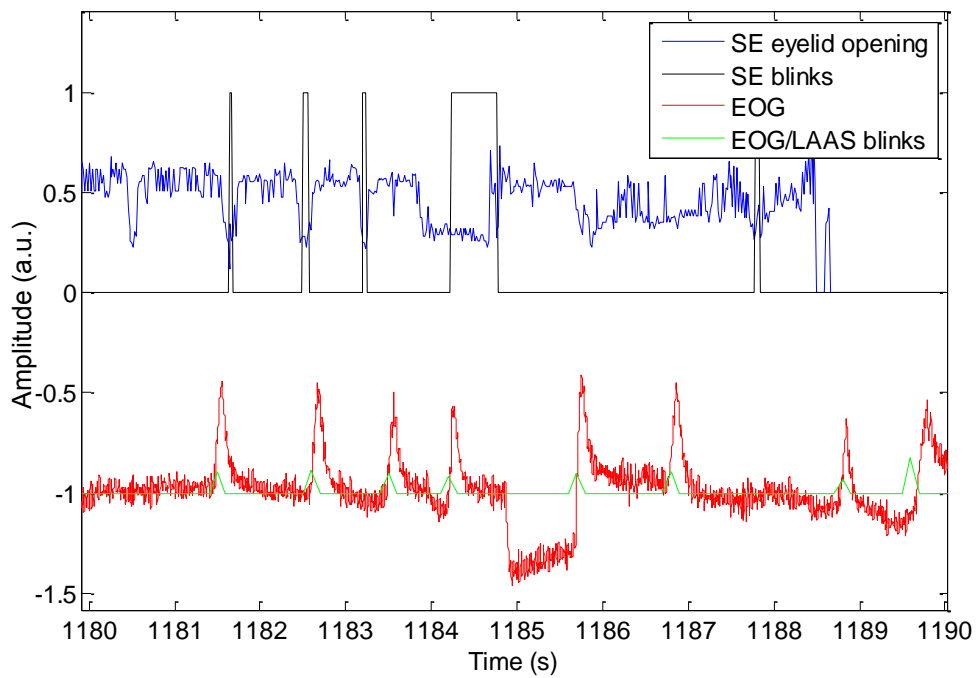


Figure 2: Number of blinks detected by LAAS and by the camera system, respectively, subdivided into day and night-time. Participants 4 and 6 cancelled their participation and are thus not included in the analysis. Data from the SleepEYE I field experiment.

In order to obtain a more accurate comparison the individual blinks identified by the two systems were compared. First, the time delay between the blink signals was determined (from the cross correlation function) and adjusted for. For each blink identified by the camera system, a matching blink in the EOG was searched for in a time window starting 0.3 s before and ending 0.3 s after the start of the (camera) blink.

When considering only matching blinks the camera system detects on average 32% of the blinks detected by LAAS (day: 32%, night: 33%). In about 2/3 of the driving sessions more than 90% of the blinks detected by the camera system could be matched to the LAAS detected blinks, which indicates that the number of erroneously detected blinks is fairly low in these sessions. However, in four driving sessions less than 60% of the blinks could be matched to the LAAS detected blinks. In one of these cases the detection rate was very low which implies that the camera system had problems detecting the blinks on the whole.

In the other three cases there were a lot of long blinks detected by the camera system. The fact that the blinks were long may explain the low rate of matching blinks (i.e. long blinks may not fit in the 0.6 s long search window). However, the reason why the blinks were long is that downward gazes were incorrectly identified as long eye closures. In another five driving sessions there were a lot of long blinks identified by the camera system that were matched to shorter blinks detected by LAAS. Also in these cases the long blinks could be explained by the fact that the camera system incorrectly had identified downward gazes as blinks (downward gazes are usually followed by a blink which explains why a shorter matching blink was found by LAAS). An example is shown in Figure 3. Another factor found to contribute to incorrect identification of long blinks is sunlight. Drivers that were exposed to sunlight tended to squint, which probably made it difficult for the camera system to decide when the blink started and ended.



*Figure 3: Example of a downward gaze that is incorrectly identified as a long blink. Note that there is approximately 1 s time delay between camera data and EOG data. Data from the SleepEYE I field experiment.*

The linear regressions between the blink durations determined by LAAS and the blink durations determined by the camera system are shown in Figure 4. In the daytime sessions the correlation between the two systems is rather low ( $r^2 = 0.13$ ,  $n = 15142$ ). In the night-time sessions the correlation is somewhat higher ( $r^2 = 0.29$ ,  $n = 13535$ ), but it is obvious that the two systems do not give similar results.

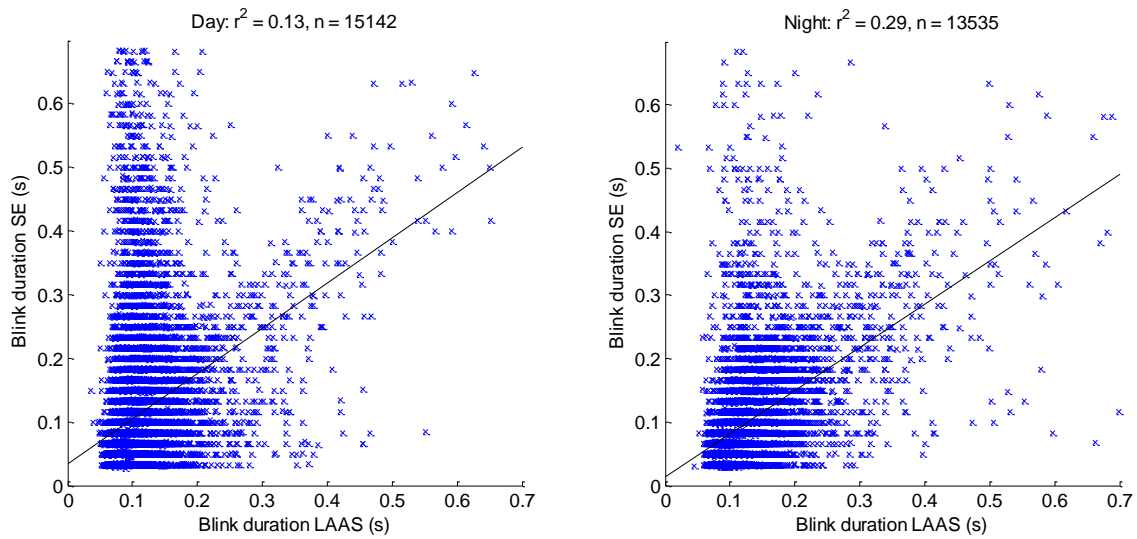


Figure 4: Linear regression between the blink durations determined by LAAS and the blink durations determined by the camera system. Left: daytime sessions, right: nighttime sessions. The discrete steps in data on the y-axis reflect the temporal resolution of the camera system. Data from the SleepEYE I field experiment.

Figure 5 shows mean blink duration obtained from the EOG and from the 1-camera system, respectively. For the EOG blinks, blink duration is higher in the night-time session than in the daytime session, which can be expected and is in line with previous research. However, for the 1-camera blinks the opposite can be seen, i.e. blink duration is higher during daytime. The result is about the same regardless of whether all blinks are included or only those that could be matched to EOG blinks. A similar result was found when the previous version of the camera system was evaluated (Fors, Ahlström et al. 2011).

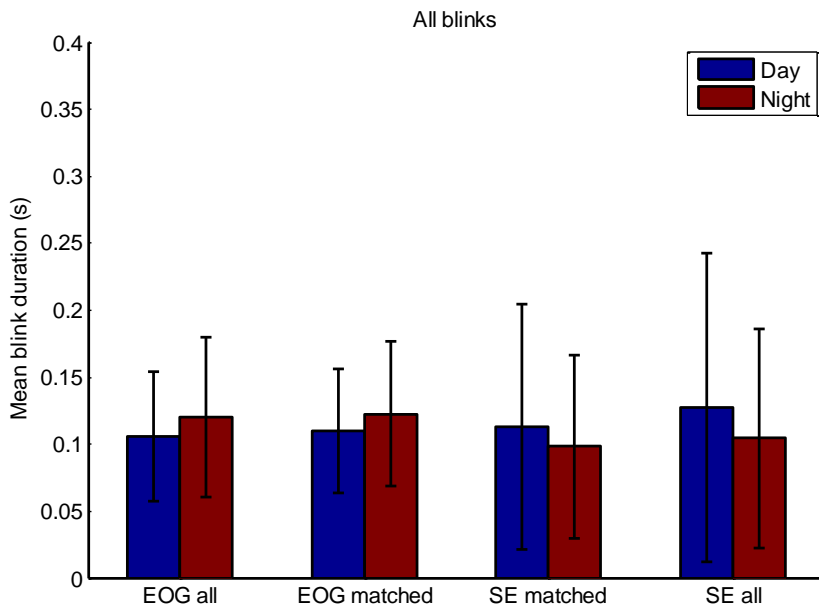


Figure 5: Mean blink duration and standard deviation for all EOG and 1-camera (SE) blinks, respectively, and for matched blinks, i.e. blinks that were identified both in the EOG and in the 1-camera data. Data from the SleepEYE I field experiment.

In fourteen out of the 36 driving sessions the detection rate of the camera system was below 25%. The eye blink data and the video films from these sessions have been visually inspected in order to identify factors that can explain the low detection rate. In nine cases the driver is looking down at the speedometer a lot (about every ten seconds or more often), which possibly could explain the poor performance of the camera system. However, there are a lot of downward gazes also in some driving sessions with a higher detection rate, which means that downward gazes do not always result in poor performance. In three of the fourteen cases with low detection rate the driver was moving his head a lot, which caused noise and/or lost tracking. In the two remaining cases there were no particular factors identified that could explain the low detection rate.

### 3.1.2 euroFOT data

Blink data quality was mainly evaluated on data from the field experiment since EOG blinks were available in that data set. However, a brief quality check was done also on blink data in the euroFOT database. Only motorway trips were included (see also Section 3.3.2).

Figure 6 shows the average number of blinks per minute per trip from euroFOT, sorted in ascending order. The blink frequency is extremely high compared to what is physiologically reasonable. As a comparison, the blink frequencies obtained by the 2-camera algorithms and by the EOG algorithm in the SleepEYE I field experiment are shown in Figure 7. The mean blink frequencies are 37 blinks/min (EOG) and 14 blinks/min (original camera algorithm), respectively. Note that the y axis scale is very different in Figure 6 and Figure 7.

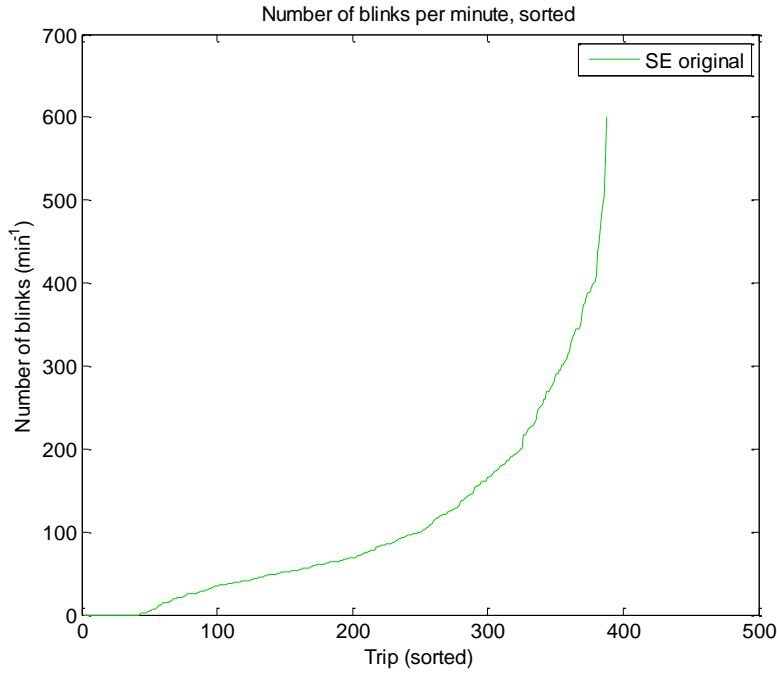


Figure 6: Average number of blinks per minute for original camera data from euroFOT, sorted in ascending order.

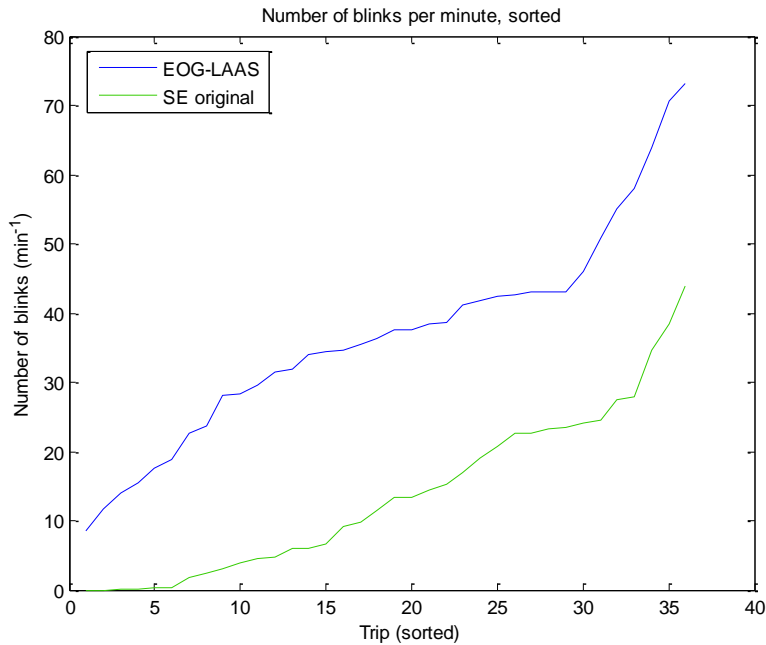


Figure 7: Average number of blinks per minute for original camera data and for EOG data from the SleepEYE I field test, sorted in ascending order.

### 3.1.3 Conclusions

To summarize, factors that were found to be associated with low blink detection rate and/or incorrect blink duration were:

- Downward gazes
- Sunlight
- Head movements

Furthermore, in euroFOT data the blink frequency was found to be far too high compared to what is physiologically reasonable and it was assumed that these data contained a lot of noise which led to false positives.

In conclusion, the evaluation of the improved version of the camera system showed that the system still missed a lot of blinks, was sensitive to noise and that there were too many incorrectly identified long blinks. The many missed blinks are unfavourable but they are perhaps not a considerable problem, as long as the blinks that are identified are correct and there are no systematic losses of data. However, the many incorrect long blinks will probably result in very poor performance when trying to identify sleepy driving in the FOT data. In addition, it was concluded that the unreasonably high blink frequency, i.e. the extremely many false positives found in euroFOT data, made it impossible to get any valid results from that dataset. Therefore, it was concluded that some post processing of the blink data was needed.

## 3.2 New blink detection algorithm

Based on the findings in the evaluation of the blink data quality a completely new blink detection algorithm was developed. This algorithm was implemented in Matlab (i.e. not in the camera system). Since the algorithm was intended to be used as a post processing algorithm for research purposes no optimization with respect to computation time had to be done, which allowed for more heavy/complex computations than what could be implemented in the (real-time) camera system.

### 3.2.1 Algorithm development

The detection of blink signals was implemented in three steps: a) the actual detection of possible blinks, b) a confirmation that the detection actually was a blink, and c) measurement of the blink duration.

Detection, confirmation and measurement are made separately for left and right eye. If a blink is confirmed for both eyes at the same time the average of both durations will be used. After a blink is detected no other blink can be detected for a period of 0.15 seconds from the opening mid time.

For detection, the Gaze Quality signal is used. If it drops to near zero it is a strong indication that a blink has occurred.

For confirmation and estimation of blink duration, template matching using cross correlation is used. The blink template can be matched on two types of signals, each with its advantages and disadvantages:

- Gaze Quality: This signal is based on the iris detector and this feature is robust due to its elliptic appearance. It is not an accurate signal of the eye opening, meaning that the blink duration measurement would be poorer.

- **Eyelid Opening:** This feature is harder to detect and is more influenced by noise and artefacts, e.g. the frame of eyeglasses. On the other hand, it can give a more accurate estimation of the blink duration.

The selected signal is then pre-filtered through a median filter of length three in order to reduce the influence of short drop-outs and outliers.

For each blink detection, a falling flank template (Figure 8, left) is matched to the signal using cross correlation. The template is varied in size to correspond to different velocities of the blink. If the match is above a given threshold the algorithm continues to check the rising flank. This is done in a similar manner, but the rising flank template (Figure 8, right) is altered to include a low value part to match the period of time when the eye stays closed. If this template then scores a correlation value over a given threshold, a blink is confirmed.

There are also some other criteria on the eyelid signal to decide whether it is a valid blink or not. These criteria were determined heuristically through testing on the dataset from the SleepEYE I field experiment. Many different criteria were tested and different versions were created. From these versions the best two candidates were selected for further testing, see Section 3.2.2.

**Algorithm 1:** The raw eyelid opening corresponding to the lowest value of the template needs to be less than 3 mm, and the corresponding highest value more than 3 mm. The difference between these two needs to be at least 2 mm, and the difference from the end point of the falling flank to the start of the rising flank has to be less than 1 mm.

**Algorithm 2:** For the Gaze Quality signal the dependences are similar, but instead of an eyelid opening in millimetres it is unit-less and divided by 10. So applying a 3 mm threshold on the eyelid opening corresponds to using 0.3 as a threshold on the gaze quality.

When a blink is confirmed the duration is determined by the distance between the closing midpoint and the opening midpoint. The midpoints are illustrated in Figure 8 as red circles.

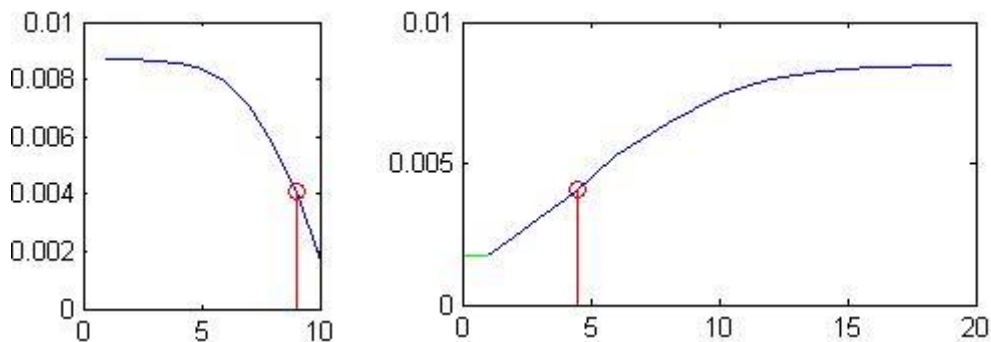


Figure 8: The falling flank template (left) and the rising flank template (right). The red circles are the closing and opening midpoints, respectively.

### 3.2.2 Final selection of blink detection algorithm

The two most promising algorithms (algorithm 1 and algorithm 2 above) were compared according to a pre-defined evaluation plan:



### ***For SleepEYE I field data***

- Describe both algorithms when applied to this dataset: pros, cons, known problems, etc.
- Number of blinks detected: EOG, matched, camera-based per driving session
- Eye tracker availability (%)
- Detection rate: camera-based compared to EOG, matched compared to EOG
  - Per driving session
  - Average day, average night
- Blink duration: EOG, camera-based
  - Mean per driving session: day vs. night
  - Mean per driving session: alert vs. sleepy<sup>2</sup>
  - Total mean: day vs. night
  - Total mean: alert vs. sleepy
  - Scatter plots of EOG vs. camera-based blink duration per driving session

### ***For euroFOT data***

*Note: Include only motorway data and only data acquired with the last version of the eye tracker software.*

- Describe both algorithms when applied to this dataset: pros, cons, known problems, etc.
- Descriptive plots:
  - Distribution of blink duration (x-axis 0–1000 ms)
  - Blink duration per trip, sorted
  - Blink duration per trip with standard deviation, sorted
  - Number of blinks per minute, sorted
  - Number of blinks per minute with standard deviation, sorted
  - Blink duration and time of day
  - Eye tracker availability (%)
- Include only trips with mean blink rate 5–75 blinks per min and generate the same descriptive plots:
  - Distribution of blink duration (x-axis 0–1000 ms)
  - Blink duration per trip, sorted
  - Blink duration per trip with standard deviation, sorted
  - Number of blinks per minute, sorted
  - Number of blinks per minute with standard deviation, sorted
  - Blink duration and time of day
  - Eye tracker availability (%)
- For trips with mean blink rate 5–75 blinks per minute, determine:
  - Number of trips
  - Number of drivers
  - Total length in minutes

---

<sup>2</sup> Sleepy is defined as KSS > 7.5

- For trips with mean blink rate 5–75 blinks per minute, check data:
  - For trips (look at several trips, different drivers) with long mean blink duration (approx.  $> 0.15$  s). Why is the blink duration long? True blinks? Systematic errors? Etc.
  - For trips (look at several trips, different drivers) with a high blink rate (approx. 60–75 blinks per minute). Why is the blink duration high? True blinks? False detections because of noise? Etc.
  - For trips (look at several trips, different drivers) with a low blink rate (approx. 5–10 blinks per minute). Why is the blink duration low? True blinks? Many misses? Any systematic misses?

The aim of the evaluation was to get enough information to be able to select the “best” algorithm. The criteria for the selection were:

- *High blink duration accuracy,*
- *No systematic losses of long blinks,*
- *No systematic false positives (especially long blinks)*
- *A detection rate that is “good enough”.*

*The duration accuracy is of greater importance than the number of detected blinks, but with a very low detection rate (<10-20%) there is a high risk that we lose too much information.*

A summary of the evaluation is shown in Table 1. In brief, algorithm 2 had a higher detection rate than algorithm 1, but algorithm 1 seemed to have a lower rate of false positives – particularly incorrectly detected long blinks – than algorithm 2. It was concluded that algorithm 1 had the best agreement with the selection criteria, and it was therefore chosen as the final blink detection algorithm. The performance of algorithm 1 is further described in next section.

*Table 1: Summary of the evaluation of the two algorithms using SleepEYE 1 field data.*

| Parameter  | Algorithm 1  | Algorithm 2  |
|--|--------------|--------------|
| Detection rate day (%)   | 54           | 81           |
| Detection rate night (%)   | 54           | 75           |
| Detection rate day, matched (%)  | 46           | 57           |
| Detection rate night, matched (%)  | 48           | 63           |
| Mean blink duration per session, number of sessions where $\text{blinkdur}_{\text{night}} > \text{blinkdur}_{\text{day}}$    | 15 out of 18 | 9 out of 18  |
| Mean blink duration per session, number of sessions where $\text{blinkdur}_{\text{sleepy}} > \text{blinkdur}_{\text{alert}}$ | 13 out of 14 | 11 out of 14 |

### 3.3 Post processed blink data

Some descriptive data of the blinks detected by the final blink detection algorithm is presented in this section.

#### 3.3.1 SleepEYE I field data

Figure 9 shows the number of blinks identified by LAAS and by the final blink detection algorithm, i.e. post processed camera blinks and the number of post processed camera blinks that could be matched to EOG blinks. On average, the new camera detection algorithm detected 54% of the blinks detected by LAAS (day: 54%, night: 54%). This is considerably higher than the corresponding figure for the original camera algorithm, which was 39% (Section 3.1). Only one of the 36 driving sessions had a detection rate less than 20%.

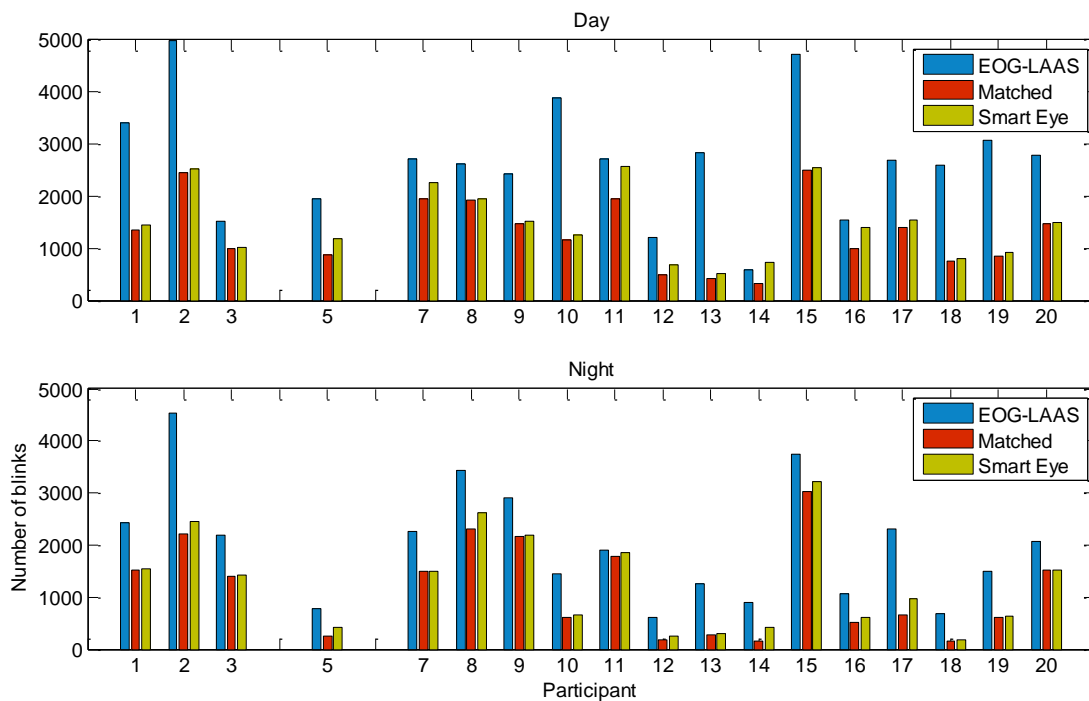


Figure 9: Number of blinks detected by LAAS and by the final blink detection algorithm i.e. post processed camera blinks (Smart Eye), respectively, and the number of post processed camera blinks that could be matched to EOG blinks. Data from the SleepEYE I field experiment.

When looking at only matched blinks the average detection rate was 47% (day: 46%, night: 48%). In about 3/5 of the driving sessions more than 90% of the blinks could be matched to EOG blinks. This is somewhat lower than for the original camera blinks, where more than 90% of the blinks could be matched to EOG blinks in about 2/3 of the driving sessions.

In two sessions less than 60% of the post processed camera blinks could be matched to EOG blinks. For the original camera blinks, a low matching rate was found to be related to incorrect identification of long blinks. For the post processed camera blinks, the problem was related to a single individual for whom several problems were present: dazzling sunlight which makes the driver to squint, noisy signals and downward gazes.

The linear regressions between the blink durations determined by LAAS and the blink durations from the post processed camera data are shown in Figure 10. The correlations are lower for the post processed blinks than for the original blinks. In the daytime sessions  $r^2$  is 0.016,  $n = 23322$  (original:  $r^2 = 0.13$ ,  $n = 15142$ ). In the night-time sessions  $r^2$  is 0.06,  $n = 20852$  (original:  $r^2 = 0.29$ ,  $n = 13535$ ). The poor correspondence between LAAS and the post processed camera blinks with respect to blink duration is a disappointing result. However, a good thing is that the great amount of long blinks incorrectly identified by the camera system, mainly in the daytime sessions, is reduced by the post processing.

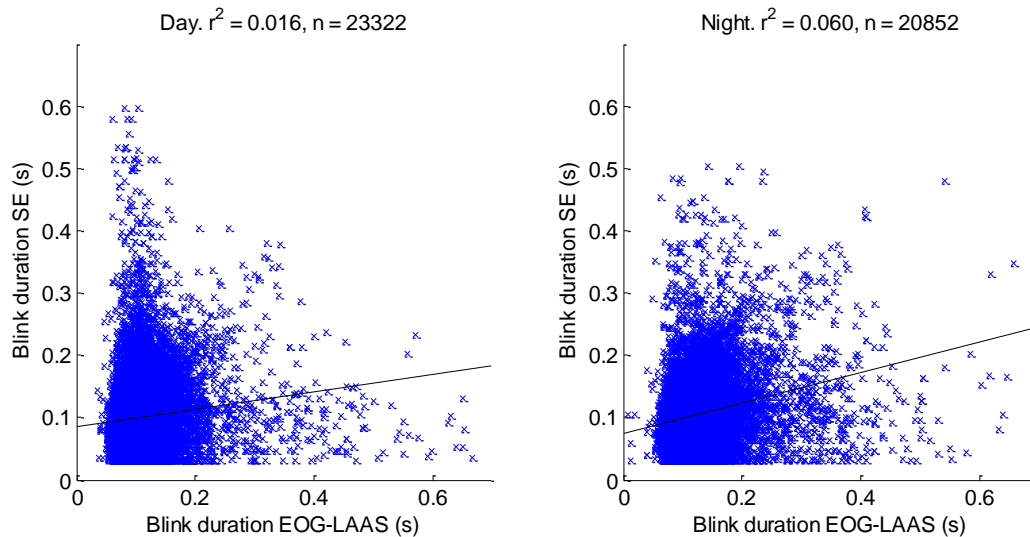
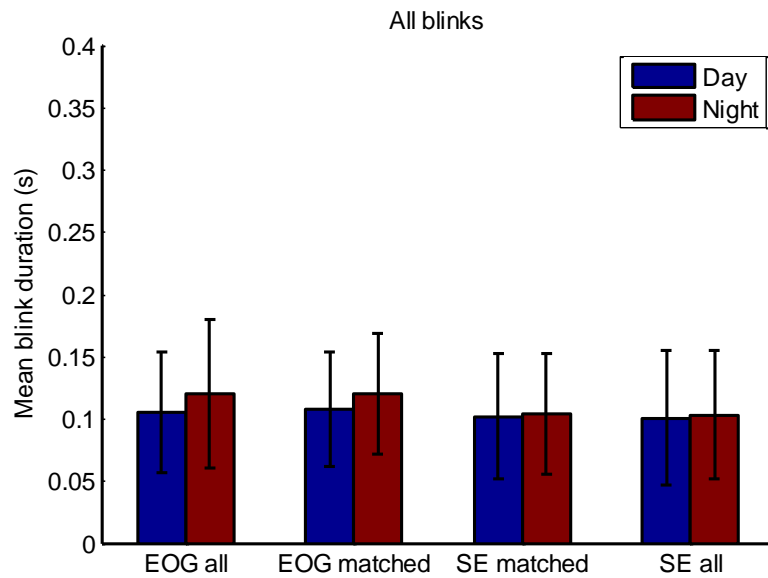


Figure 10: Linear regression between the EOG-LAAS blink durations and the blink durations from the final version of the post processed camera data (SE). Left: daytime sessions, right: night-time sessions. Data from the SleepEYE I field experiment.

The reduction of incorrectly identified long blinks is also reflected when comparing daytime blink duration with night-time blink duration, Figure 11. For the original camera data blink duration was lower at night than in the daytime sessions (Figure 5), but for the post processed camera data the opposite was found, which is the expected result and also in agreement with the LAAS blinks. When looking at individual driving sessions, the mean duration of LAAS blinks is higher at night than during the day for all 18 participants. For the post processed camera blinks the mean blink duration is higher at night than during the day for 15 out of the 18 participants.



*Figure 11: Mean blink duration and standard deviation for all EOG and post processed camera (SE) blinks, respectively, and for matched blinks, i.e. blinks that were identified both in the EOG and in the post processed camera data. Grouped by driving session. Data from the SleepEYE I field experiment.*

If blink durations are grouped into alert ( $KSS \leq 7.5$ ) and sleepy ( $KSS > 7.5$ ) the results look even more promising, Figure 12. When looking at individual driving sessions, the mean duration of LAAS blinks is higher when sleepy than when alert for all 15 participants that became sleepy. For the post processed camera blinks the mean blink duration is higher when sleepy than when alert for 14 out of the 15 participants.

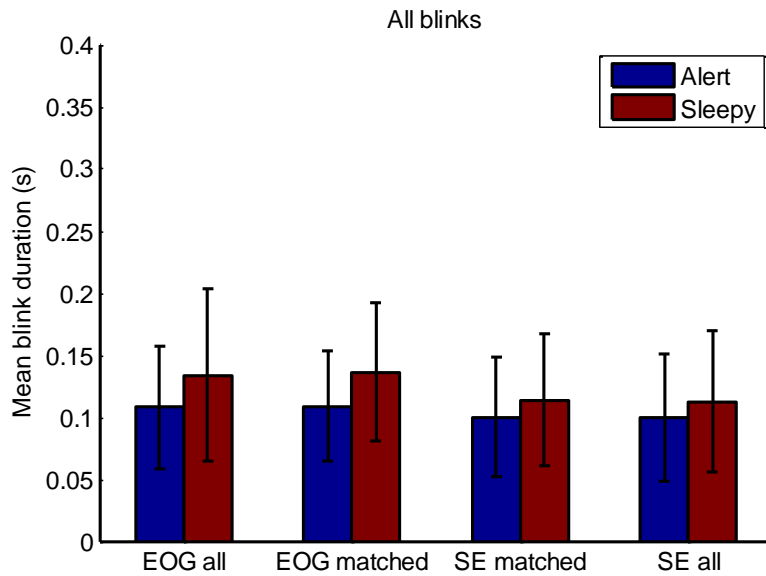


Figure 12: Mean blink duration and standard deviation for all EOG and post processed camera (SE) blinks, respectively, and for matched blinks, i.e. blinks that were identified in both the EOG and in the post processed camera data. Grouped by KSS. Data from the SleepEYE I field experiment.

Figure 13 shows mean blink frequency per session, sorted by LAAS blink frequency in ascending order. For LAAS blinks the average blink duration is 37 blinks/min and it varies from 9 blinks/min to 73 blinks/min, while the blink duration of camera blinks range from 5 to 47 blinks/min with an average of 21 blinks/min. It is obvious that there is a lack of relationship between the two algorithms, with respect to blink frequency. However, blink frequency is not very interesting as a sleepiness indicator, so this might not be a problem. What also can be seen is that low blink frequencies may indicate a low detection rate. Figure 14 shows the relationship between detection rate and blink frequency for the post processed camera blinks. Blink frequency tends to increase with the detection rate, and it may thus be possible to use blink frequency as a quality criterion. Based on Figure 13 and Figure 14, it is suggested that a blink frequency of 5–75 blinks/min should be regarded as valid.

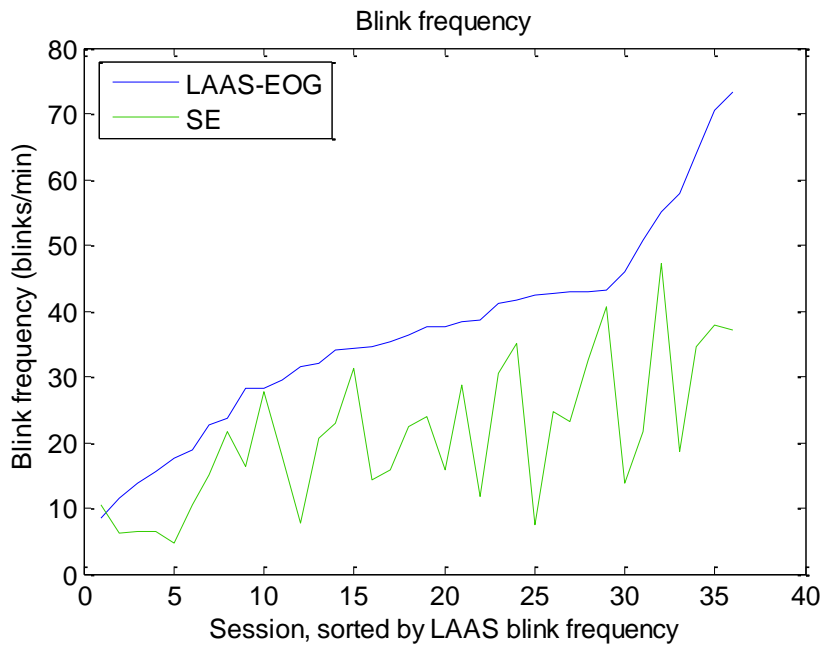


Figure 13: Blink frequency per session, for LAAS and post processed camera blinks, sorted by LAAS blink frequency in ascending order. Data from the SleepEYE I field experiment.

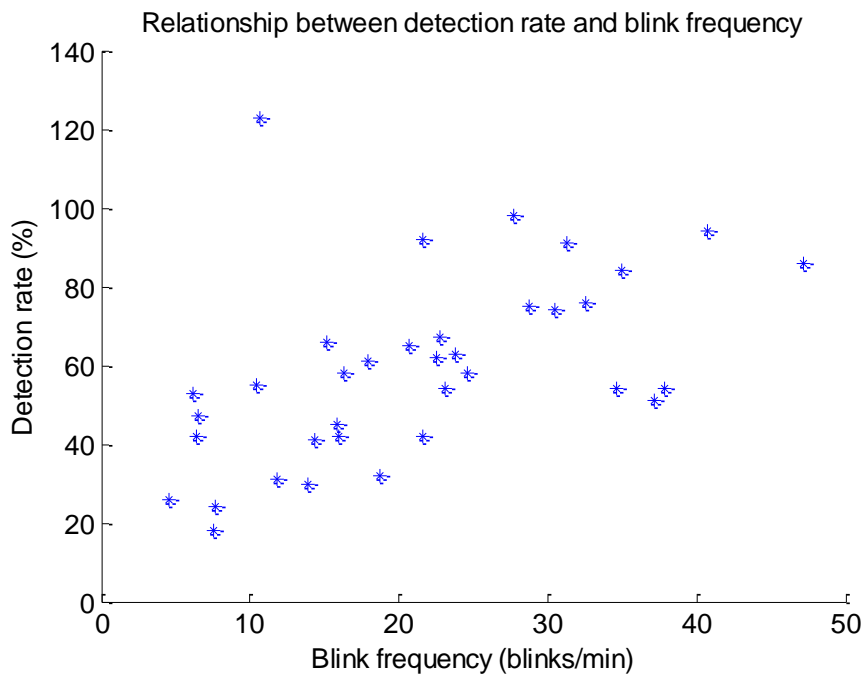


Figure 14: Relationship between detection rate and blink frequency for the post processed camera blinks. Data from the SleepEYE I field experiment.

In conclusion, the final blink detection algorithm identifies about half of the blinks identified by LAAS. Blink duration shows a poor correspondence for individual blinks,

but it looks relatively good on the group level. No systematic errors have been identified. It is suggested to use blink rate as a quality criterion, and to exclude driving sessions that have a blink rate outside the range of 5–75 blinks/min.

### 3.3.2 euroFOT data

All trips that had the latest version of the camera software and that were undertaken on motorways were filtered out from the euroFOT database. In total 2442 trips had the latest version of camera software, 715 of these trips did not have map data (thus it was not possible to calculate the type of road) and 192 trips did not have necessary signals to calculate the blink duration. Out of the remaining 1535 trips there were segments driven on motorway in 388 trips. Thus, in total 388 trips were filtered out, and from these trips only the motorway segments were included in the analysis. Eye tracking data from this subset of trips were run through the new blink detection algorithm (described in Section 3.2) and the results were summarized in the graphs below.

Figure 15 shows the average number of blinks per minute per trip. In 208 trips the blink frequency is very low ( $< 5$  blinks/min), which probably indicates that the algorithm has problems detecting blinks. There are no trips that have an unreasonably high blink frequency, in contrast to the original blink data obtained from the camera software (Figure 6), which indicates that the new blink detection algorithm has a relatively low rate of false positives (i.e. noise classified as blinks).

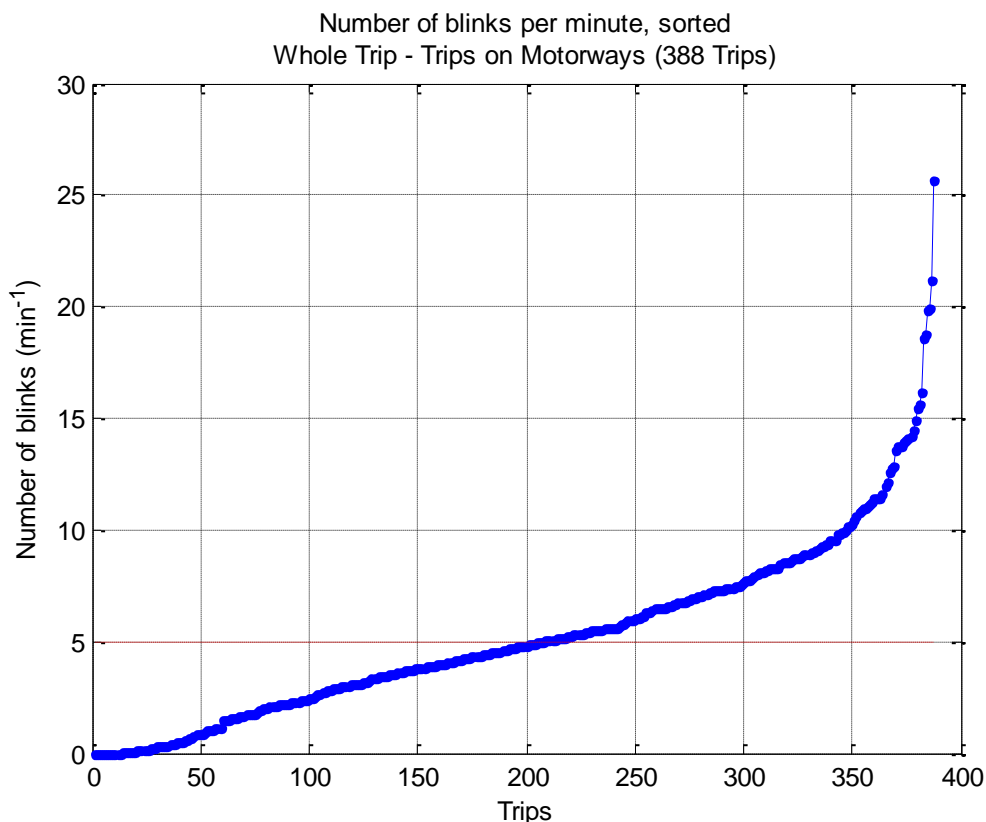


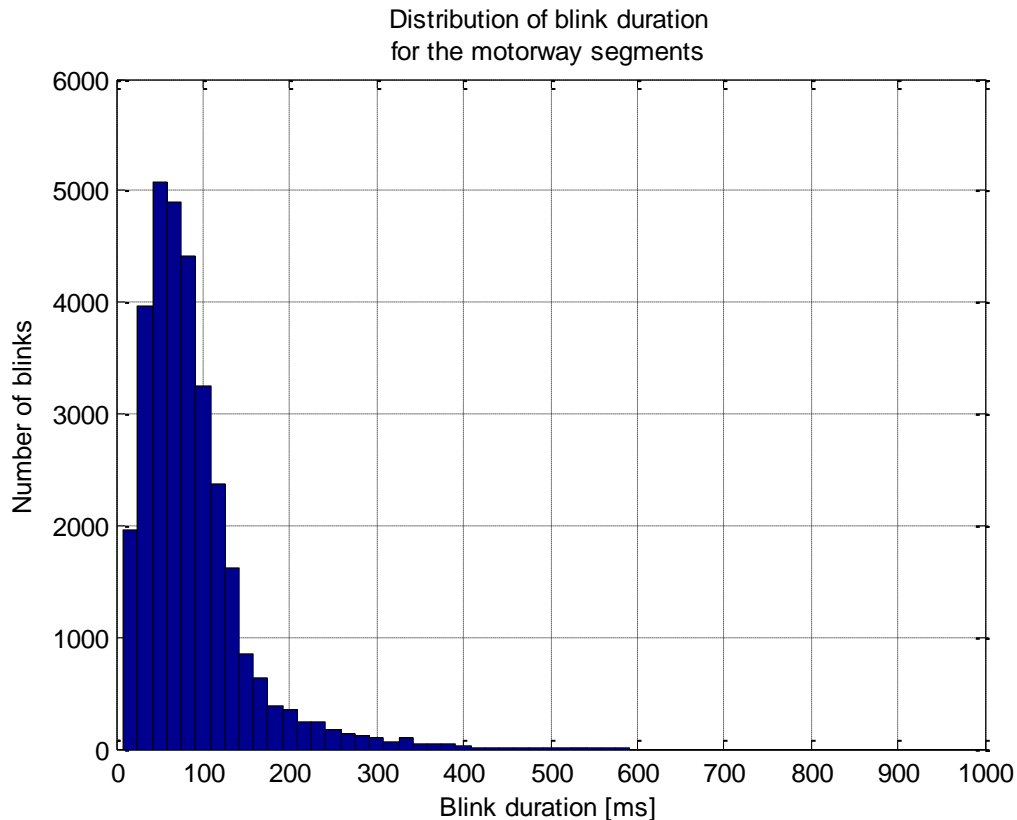
Figure 15: Number of blinks/min per trip (blue), sorted by number of blinks/min. The red line corresponds to 5 blinks/min. Data from euroFOT.

In order to keep only the most reliable data for analysis all trips where the mean blink frequency was outside the interval 5–75 blinks/min were excluded, as suggested in



Section 3.3.1. For the remaining 180 trips the analysis was done only for the segments of the trips that were driven on motorways. Figure 16 and Figure 17 show descriptive data for the trips where the mean blink frequency was 5–75 blinks/min.

Figure 16 shows the distribution of blink duration. The shape of the distribution corresponds to what could be expected: most blinks are found in a relatively narrow duration interval, while a small amount of the blinks have a longer duration.



*Figure 16: Distribution of blink duration from motorway segments. Trips where mean blink frequency is outside the interval of 5–75 blinks/min are excluded. Data from euroFOT.*

Figure 17 shows mean blink duration and standard deviation per trip, sorted by blink duration. It is impossible to get any detailed information about the performance of the blink detection algorithm based on this figure, but what can be said is that there are no abnormalities identified in these data. However, it can be seen that there are very large variations in blink duration within the trips, but the reason for this is not known.

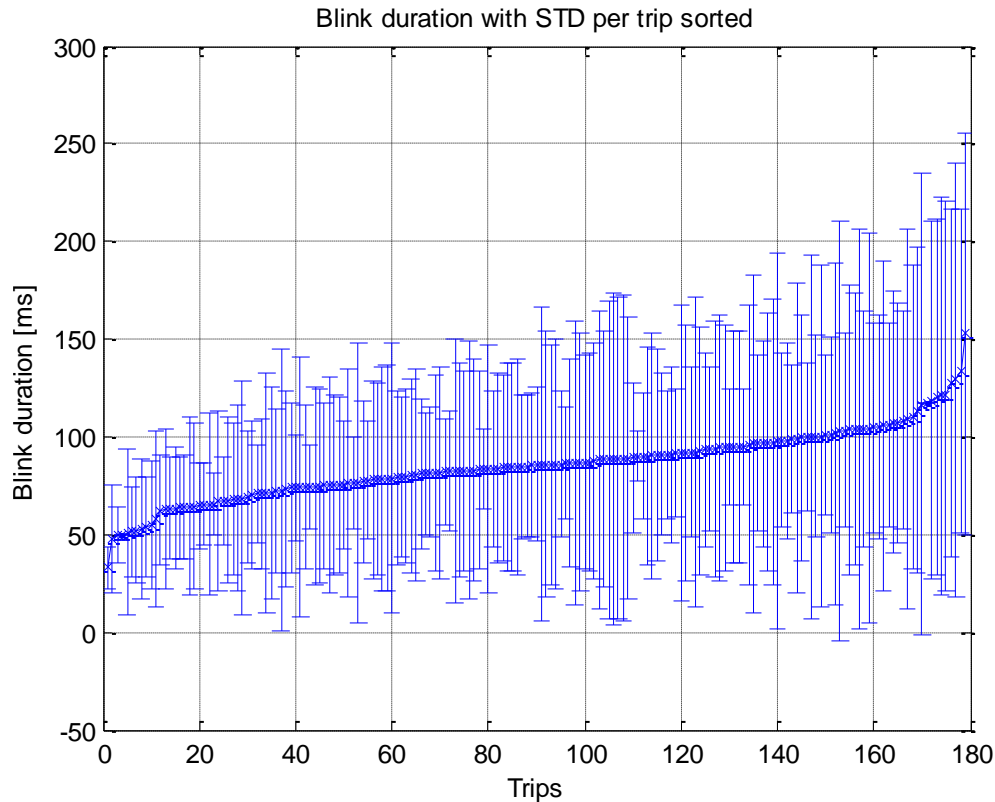


Figure 17: Blink duration (mean and standard deviation) per trip (motorway only) for blink frequencies in the interval 5–75 blinks/min, sorted by blink duration. Data from euroFOT.

### 3.4 Summary and conclusions

The improved camera software was evaluated and it was found that it missed a lot of blinks, was sensitive to noise, and incorrectly classified downward gazes as long blinks. Therefore, it was concluded that some post processing of the blink data was needed in order to further improve blink data quality. This work was done in several steps and it finally ended up in a completely new blink detection algorithm which was implemented in Matlab (i.e. not in the camera software). The new algorithm had a better detection rate and a lower rate of false positives than the algorithm implemented in the camera system, and it was decided that this was the final algorithm that should be used in the project.

It was recommended that only trips with a mean blink frequency within the range 5–75 blinks/min should be included in the analyses, in order to minimize the number of trips where problems with the blink detection can be suspected.

## 4 Sleepiness indicators: selection, performance and analysis

The sleepiness detector developed in SleepEYE I, which was intended to be used in driving simulators, was based on the following sleepiness indicators:

- *fracBlinks*: fraction of blinks  $> 0.15$  s, either from a camera-based system or from the electrooculogram (EOG)
- *SDLP*: standard deviation of lateral position
- *KSS<sub>estSR</sub>*: an estimated sleepiness level, based on the sleep/wake predictor (Åkerstedt, Folkard et al. 2004) and the KSS value (Åkerstedt & Gillberg 1990) reported by the driver immediately before the driving session

The indicators were calculated over a sliding window of 5 minutes. The detector aimed to classify whether the driver reported a KSS value of 8 or higher (actually higher than 7.5, since the sliding window approach may result in a value with decimals), where 8 corresponds to “*sleepy and some effort to stay awake*”. A set of decision rules, in the form of a decision tree, was generated using the open source algorithm C5.0<sup>3</sup>.

The selection of indicators was made with the simulator setting in mind. Since the project had a primary focus on camera-based indicators such indicators were naturally included in the development process. In particular blink related parameters were found to be very promising for sleepiness detection, based on the literature review that was carried out in the beginning of the project. In the literature review support was also found for using model based information (i.e. predictions of sleepiness level) and variations in lateral position as sleepiness indicators. In a previous study model based information, i.e. a sleep/wake predictor (SWP), alone was found to be able to detect sleepiness with a performance score of 0.78 (Sandberg, Åkerstedt et al. 2011). The SWP needs information about prior sleep, time awake and time of day, and since this information usually is available in controlled experiments it was assumed to be a feasible indicator. Variation in lateral position is often used as a general measure of driving performance and it has been found to be related to sleepiness in several previous simulator studies (Liu, Hosking et al. 2009).

In the present project the aim was to adapt the sleepiness indicators described above to a field setting and ultimately to be applicable to FOT data.

### 4.1 SleepEYE I field data

#### 4.1.1 Selection of indicators for field data

In a first step, the sleepiness indicators that were used in SleepEYE I were applied to data from the field experiment. For each single indicator the optimal threshold with respect to performance – defined as the mean of sensitivity and specificity – was calculated, as described in the SleepEYE I report (Fors, Ahlström et al. 2011).

All indicators were calculated in time windows of 5 min with 1 min resolution. The target was to separate  $KSS > 7.5$  from  $KSS \leq 7.5$ .

The results were compared to the corresponding results from the driving simulator study.

---

<sup>3</sup> Available at [www.rulequest.com](http://www.rulequest.com)

### Blink based indicators

Four variants of blink based indicators were assessed in SleepEYE I: mean blink duration, median blink duration, mean of 25% longest blinks and fraction of blinks longer than 0.15 s. The thresholds and performances of these indicators are shown in Table 2. As a comparison, the corresponding figures for the driving simulator data are included in the same table. It should be noted that the camera-based indicators for the field test are based on the new blink detection algorithm (post processed), while the simulator counterpart is based on the old camera software.

For the simulator test, the performances of the EOG-based and the camera-based indicators were similar, with a slight advantage for the camera-based indicators. For the field test, the performance of the EOG-based indicators was comparable to that of the simulator test, while the performance of the camera-based indicators was somewhat worse.

*Table 2: Thresholds and performances of the blink based indicators. Note that the camera-based indicators from the simulator test originate from another algorithm than those from the field test.*

| Measure                            | Field     |             | Simulator |             |
|------------------------------------|-----------|-------------|-----------|-------------|
|                                    | Threshold | Performance | Threshold | Performance |
| EOG mean blink duration            | 0.11 s    | 0.70        | 0.17 s    | 0.65        |
| EOG median blink duration          | 0.11 s    | 0.72        | 0.13 s    | 0.64        |
| EOG mean of 25% longest blinks     | 0.19 s    | 0.68        | 0.23 s    | 0.71        |
| EOG fraction of blinks > 0.15 s    | 0.19      | 0.67        | 0.26      | 0.68        |
| Camera mean blink duration         | 0.09 s    | 0.61        | 0.11 s    | 0.71        |
| Camera median blink duration       | 0.10 s    | 0.61        | 0.10 s    | 0.70        |
| Camera mean of 25% longest blinks  | 0.14 s    | 0.60        | 0.27 s    | 0.69        |
| Camera fraction of blinks > 0.15 s | 0.08      | 0.60        | 0.10      | 0.72        |

### Lateral position variation

The threshold and performance of *SDLP* is shown in Table 3. The optimal threshold in the field test was identical to that in the simulator test. However, the performance was worse for the field test. Number of (unintentional) line crossings had a similar performance in the field test as in the driving simulator test.

*SDLP* and number of line crossings were calculated in sliding time windows of 5 min with 1 min resolution (as all other indicators, as described in Section 4.1.1). *SDLP* was only calculated in time windows where the vehicle was in the right lane. Road segments where the lane tracker data was of poor quality or where the driver changed lane or drove in the left lane were excluded from the time windows. If more than 20% of the data in a time window had to be excluded due to the reasons above the entire time window was excluded from the analysis.

Table 3: Threshold and performance of SDLP in the field experiment and the simulator experiment, respectively.

| Measure                                 | Field     |             | Simulator |             |
|---|-----------|-------------|-----------|-------------|
|   | Threshold | Performance | Threshold | Performance |
| Std of lateral position ( <i>SDLP</i> ) | 0.28 m    | 0.63        | 0.28 m    | 0.73        |
| Number of line crossings (per 5 min)    | 0.09      | 0.67        | 0.90      | 0.69        |

### Model based information

In SleepEYE I the SWP was used in combination with a simple model of how KSS changes with trip duration (time on task). By fitting a linear function to the KSS values reported every five minutes during the driving simulator sessions (averaged over all participants and conditions), and adding the slope of the function to the KSS value predicted by the SWP, the following model was obtained:

$$KSS_{estSWP\_sim} = 0.044t + KSS_{SWP}.$$

where  $KSS_{SWP}$  is the KSS value predicted by the SWP at the start of the driving session and  $t$  is time in minutes from the start of the driving session. The coefficient 0.044 is the slope of the linear function.

When developing the classifier in SleepEYE I  $KSS_{SWP}$  was replaced by  $KSS_{SR}$ , which is the self-reported KSS value just before the start of the driving session. The reason for doing that was that using  $KSS_{SWP}$  resulted in a very poor performance for the validation data set, which included data from severely sleep deprived participants (which therefore did not correspond to the expected SWP pattern). In SleepEYE II no self-reported KSS values are available from the euroFOT data. Therefore, using  $KSS_{SR}$  is not an option. On the other hand, in order to use  $KSS_{SWP}$  information about the drivers' sleep and wake up times is needed. Neither is this information available from the euroFOT data. A possible solution would be to estimate sleep and wake up times. A simple way of doing that is to assume that the drivers always go to bed at a certain time and that they always get up at a certain time. From here, this simplified SWP is denoted SSWP and the KSS values predicted from this model are denoted  $KSS_{SSWP}$ . Using a similar approach as in SleepEYE I, KSS for field data can then be estimated according to:

$$KSS_{estSSWP} = kt + KSS_{SSWP},$$

where  $k$  is the change in KSS per minute, i.e. the slope of the curve.  $k$  was determined from the KSS curves from the field experiment, Figure 18. Separate functions were fitted to the daytime and the night-time data:

$$KSS_{esttripdurDay} = 0.025t + 3.66$$

$$KSS_{esttripdurNight} = 0.027t + 6.00$$

Taking the average of the daytime and night-time slopes resulted in the following model:

$$KSS_{estSSWP} = 0.026t + KSS_{SSWP}$$

In order to avoid invalid KSS values the following modification was done:

$$KSS_{estSSWP} = \begin{cases} KSS_{estSSWP}, & \text{when } KSS_{estSSWP} \leq 9 \\ 9, & \text{when } KSS_{estSSWP} > 9 \end{cases}$$

When analysing data from the field experiment it is possible to use the original SWP instead of the simplified one, since sleep and wake up times are known.

Table 4 shows the thresholds and the performances of the two KSS estimates.

In the SSWP, sleep and wake up times were assumed to be 06:30 and 22:30, respectively.

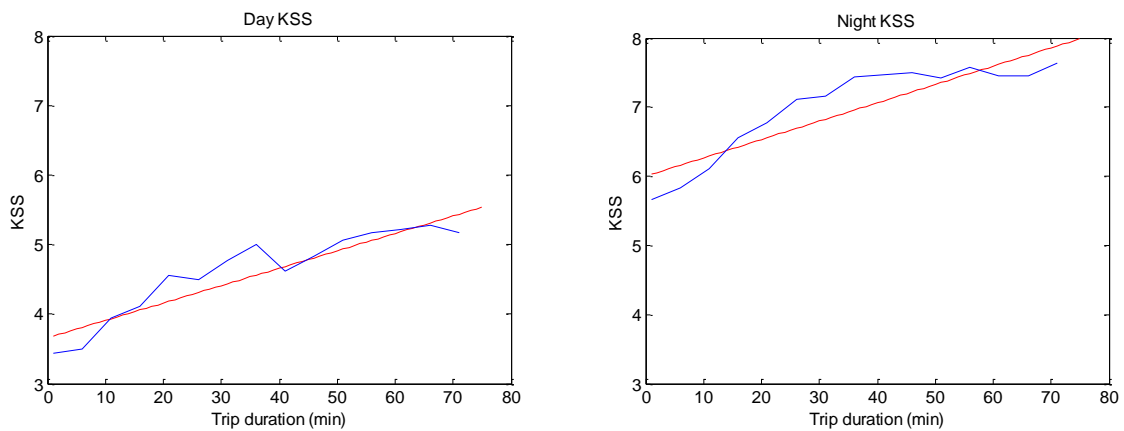


Figure 18: Mean KSS (blue) for all participants in the SleepEYE I field experiment and the linear functions that were fitted to the KSS curves (red). Left: daytime session, right: night-time session. A KSS values of 3 corresponds to "alert" while 8 corresponds to "sleepy, some effort to keep awake".

Table 4: Thresholds and performances for the KSS estimates when applied to data from the field experiment and from the simulator experiment.

| Measure         | Field     |             | Simulator |             |
|-----------------|-----------|-------------|-----------|-------------|
|                 | Threshold | Performance | Threshold | Performance |
| $KSS_{estSWP}$  | 6.99      | 0.78        | 7.95      | 0.82        |
| $KSS_{estSSWP}$ | 6.83      | 0.78        | -         | -           |

#### 4.1.2 Discussion and recommendations

The camera-based indicators and *SDLP* had a worse performance when applied to field data than when applied to simulator data. There are several possible explanations for this fact. First, the driving simulator provides an experimental setting where things can be kept constant, e.g. traffic density, light conditions and road properties. Thus, indicators that are sensitive to external non-constant factors, such as lateral position variation that is influenced by for example other traffic and road condition, can be expected to have a poor performance on real roads. Second, data quality is usually worse in real road tests compared to driving simulator tests, because of sensor failure, poor accuracy or noise. Another reason for the poor indicator performance in the field test is that the participants became less sleepy than in the driving simulator test, which is why the signs of sleepiness may be less pronounced. A related issue is that the perceived risk is greater on real roads, which may result in a difference in driving behaviour between the two tests.

Because of the poor performance, neither the camera-based blink indicators nor *SDLP* seem to be very feasible sleepiness indicators for field applications. It is a bit unexpected that the camera blink indicators have such a poor performance, when almost all participants had a higher blink duration at night than during the day, and also a higher blink duration when sleepy than when alert (see Section 3.1.1). Perhaps there are still some problems with the blink detection that were not identified in the evaluation phase.

The EOG-based indicators had a better performance than the camera-based counterpart and they may thus have a greater potential to be used in field tests. The performance is however only moderate, so these indicators probably need to be combined with some other indicators in order to provide any useful results.

The number of (unintentional) line crossings had a somewhat better performance than *SDLP* and is thus probably the best driving related indicator. Similar to *SDLP*, the number of line crossings will be influenced by external factors, but probably not to the same extent as *SDLP*.

The *KSS* estimates showed relatively good performances, which were similar to the results from the simulator. It should, however, be noted that these indicators have not been evaluated on any other dataset, and that the assumed sleep and wake up times (which are needed for the calculation of  $KSS_{estSSWP}$ ) are very similar to the real sleep and wake up times in this dataset. Furthermore, when using  $KSS_{estSSWP}$  individual variations are not taken into account at all. The indicator will only depend on time of day.  $KSS_{estSSWP}$  could perhaps be used in controlled tests where the participants' sleep and wake up times are known.

In conclusion,  $KSS_{estSSWP}$ , number of line crossings, and EOG-based indicators are the most promising indicators for controlled field tests, but they are probably not very useful as individual indicators. Some data fusion and decision algorithms will be needed, and probably also additional indicators.

## 4.2 From SleepEYE field data to euroFOT data

### 4.2.1 Selection of indicators for euroFOT data

As demonstrated in the previous section, finding robust sleepiness indicators for field applications is very difficult and regardless of how the suggested indicators are selected

or combined (e.g. fed into a classifier) the performance can be expected to be only moderate, in the best case. FOT applications put further limitations on what can be achieved. Usually less is known about the context (e.g. road characteristics) and about relevant background variables (e.g. hours slept). With that said, it is still interesting to assess the applicability of the suggested indicators on the euroFOT database, in order to take a small step towards a methodological framework for sleepiness analysis in FOT data.

The most promising indicators for field data were  $KSS_{estSWP}$ , number of line crossings, and EOG-based indicators. Some modifications need to be done before applying these indicators to euroFOT data:

- $KSS_{estSWP}$  must be replaced by  $KSS_{estSSWP}$  since no information about the drivers' sleep and wake up times is available in euroFOT. But, by using  $KSS_{estSSWP}$  individual variations in sleep pressure will be hidden, and there is a risk that this indicator will make it more difficult to identify sleepy drivers, particularly in the euroFOT dataset where the vast majority of trips are undertaken during daytime (see Chapter 6). Thus, it was decided to skip KSS estimates as indicators for the euroFOT dataset.
- EOG-based indicators must be replaced by the camera-based counterparts since no EOG data is available in euroFOT

The performances of the four camera-based indicators were similar and it is likely that these indicators contain basically the same information. Therefore, it was decided that only one of these indicators should be included. Thus, the two indicators suggested to be used for assessment of driver sleepiness in euroFOT were:

- Mean blink duration (camera-based)
- Number of line crossings

#### 4.2.2 Application of euroFOT indicators on SleepEYE data

In order to get a better understanding of the indicators and their relationship to each other and to the level of sleepiness, but also as an input to the decision of how to use the indicators in euroFOT, an explorative analysis of the data from the SleepEYE I field test was done.

All indicators were calculated in time windows of 5 min with 1 min resolution. The data points (cases) were labelled either as alert ( $KSS \leq 7.5$ ) or sleepy ( $KSS > 7.5$ ).

Figure 19 shows a scatter plot of number of line crossings and mean blink duration (camera-based). It is obvious that there is a great overlap between alert and sleepy cases, and that it is impossible to separate the vast majority of the cases. However, when looking only at alert cases, it can be seen that most of them can be found in the lower left part of the figure. A simple classification function can be drawn around this area just by looking at the figure (the blue line in Figure 19). The performance of this classifier is shown in Table 5. Only 2.1% of the alert cases, corresponding to 39 cases, are incorrectly classified, which is an interesting result.

Table 5: Performance of the classifier defined by the blue line in Figure 19

|  | Classified as | Performance |
|--|---------------|-------------|
|--|---------------|-------------|



| Observations | Alert | Sleepy | % correct | % incorrect |
|--------------|-------|--------|-----------|-------------|
| Alert        | 1814  | 39     | 97.9      | 2.1         |
| Sleepy       | 257   | 107    | 29.4      | 70.6        |

By analysing the 39 alert cases, which were incorrectly classified as sleepy, the following was found:

- 25 cases had a KSS  $\geq 6$ .
- 6 cases were from a night-time session, after 30-40 min of driving, where the driver reports KSS 5.
- 4 cases were from a daytime session, where the driver reports KSS 5-6. 5–10 min later, this driver reports KSS 7. The average KSS for the entire session is approximately 7, i.e. this driver is very sleepy.
- 4 cases were from a daytime session, where the driver reports KSS 3. It is believed that this driver had misunderstood the KSS scale. The test leader in the car reported that the driver was sleepier than the ratings, and from the video recordings it can be seen that this driver has periods of long blink durations.

Thus, there is some evidence that the drivers actually were sleepy in many of the incorrectly classified 39 cases.

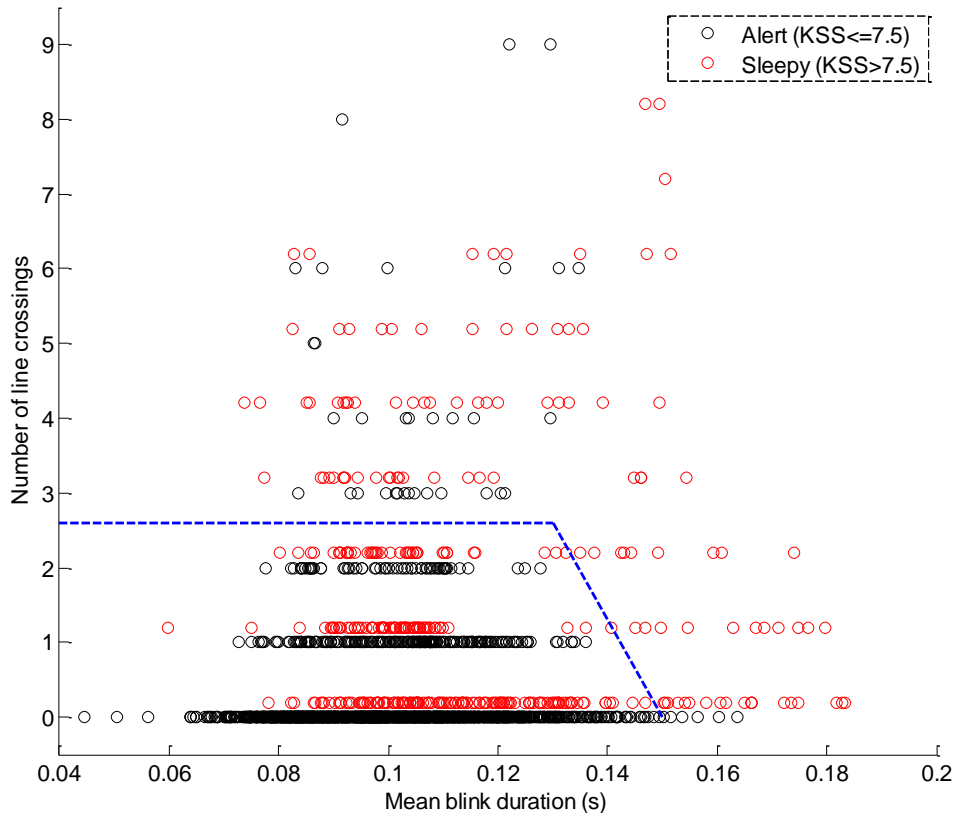


Figure 19: Scatter plot of number of line crossings and mean blink duration (camera-based) for alert (black) and sleepy (red) cases in the SleepEYE I field dataset. The blue line indicates an example of a classification function. A small offset is added to the number of line crossings for the sleepy cases, in order to make both alert and sleepy cases visible.

Sleepy driving is harder to classify than alert driving. Only about 30% are correctly classified with the simple classifier in Figure 19. An analysis of the correctly classified 107 cases shows that they represent 8 out of 14 drivers that became sleepy, which is a very positive finding. If the correctly classified sleepy cases had originated from only a single or a few individuals the classifier had been more or less useless. It should however be noted that all but two correctly classified sleepy cases with a blink duration > 0.14 s originate from one individual (approx. 50 cases).

For 6 drivers none of the sleepy cases were correctly classified. For 7 drivers there were both correct and incorrect classifications. Thus, the incorrectly classified sleepy cases represent 13 out of the 14 drivers that became sleepy. For one driver, all sleepy cases were correctly classified.

This fairly simple example demonstrates that it should be possible to construct a classifier that identifies sleepy driving with a low sensitivity but with a high specificity. Although a high sensitivity also is desired, this classifier will allow for a first rough assessment of sleepiness in the euroFOT data.

## 5 Video ORS

As noted in the introduction above, the most common approaches to studying driver sleepiness is considering subjective estimations, physiology or driving performance (related to sleepiness). There is, however, another possible approach to measuring driver sleepiness, which has gained little attention in the literature, namely **Observer-Rated Sleepiness (ORS)**. The idea is simply to let an observer estimate the level of sleepiness in the test subject by *observing* the test subject. The idea is based on the fact that sleepiness has an effect on behavioural functionality with, in most cases, apparent behavioural signs of sleepiness such as yawning, nodding.

ORS estimations may be carried out in different ways. In a driver sleepiness study, where there are test leaders in the car together with the test subject, the ORS estimations may be carried out in real time by the test leader, whereas in a large scale naturalistic study, where there are no test leaders, the ORS estimations can be done using video recordings of the drivers.

As mentioned above there has been very minimal methodological research on ORS in the literature on driver sleepiness, with the work of Wierwille and Ellsworth (1994) being the only work known to the authors. Still, ORS has been employed both in large scale naturalistic studies, such as the 100-car study (Dingus, Neale et al. 2006) to draw conclusions on the increased risk of accidents for sleepy drivers, and in work on developing driver sleepiness detection and warning systems.

The idea to be able to estimate the level of driver sleepiness using video recordings of the driver is an attractive idea in large scale naturalistic studies where no other measures of sleepiness are recorded. (It should be noted that it may be possible to infer driver sleepiness based on driving behaviour based measures, although such measures are likely to be very context-dependent and contain much noise.)

In the work by Wierwille and Ellsworth (1994) six graduate students (from a program involving human factors methodology) were recruited to participate in the study. 48 one-minute video clips of drivers' faces were seen by the participants who were tasked with estimating the sleepiness of the drivers in the video clips using a Likert scale. The results showed a good intra-rater and inter-rater reliability and good validity when compared to other measures of sleepiness (such as eyelid movements). However, very little details are given in the paper. For example, details on video quality, experimental setup of the study in which the videos were recorded, etc. are few or missing.

Furthermore, the videos were recorded in a simulator study, which ensures much better video quality than what may be obtainable in a large scale naturalistic driving setting. Finally, the raters' estimations were not primarily compared with some ground-truth of sleepiness but rather with the average rated estimation for each clip. Some work on validation is described briefly in the discussion section of the paper. Although details are brief, there seems to be some validity of the raters' estimations with other measures of sleepiness (please see the paper for more details).

This motivated the methodological research on validating video-based ORS estimations that was carried out within the framework of SleepEYE II.

### 5.1 Method

Forty participants were recruited to take part in the study as *raters*. The basic idea of the study was to let the raters watch video clips of drivers who took part in the SleepEYE I

field study, which was carried out in 2010, and estimate the drivers' levels of sleepiness. The raters were also tasked with giving their level of confidence (in steps of 10% on a scale from 0%, meaning no idea, to 100%, meaning being certain) for each sleepiness estimation made. A set of 54 one-minute video clips were selected from the SleepEYE I field study. The raters watched and scored these 54 clips in three sessions where 18 clips were shown in each session.

The problem is of course how to validate the sleepiness ratings given by the raters towards some *true* value of how sleepy the drivers in the video recordings were. Here it was decided to use the drivers' own subjective sleepiness estimations on the Karolinska Sleepiness Scale (KSS) to define the true value of sleepiness in each video clip. KSS ratings have been shown to correspond with physiological signs of sleepiness and have been extensively validated in the literature (Åkerstedt & Gillberg 1990; Kaida, Takahashi et al. 2006). The drivers in the SleepEye I field study estimated their subjective level of sleepiness, by answering the question "How sleepy have you felt during the last five minutes", every 5<sup>th</sup> minute while driving. It was not (and should not) be expected that subjective sleepiness and behavioural signs of sleepiness will correlate perfectly. On the other hand, some measure of sleepiness must be used as the reference and subjective sleepiness measures provide the best reference we have available in a driving context today.

The raters were tasked with estimating the driver's sleepiness level in each video clip on a three graded scale, namely the ORS. The ORS scale has been under development during the last few years by researchers at VTI and the Stress research institute at Stockholm University (see (Anund, Fors et al. 2013) for further details).

Since the KSS scale has nine levels whereas the ORS has only three levels a mapping of ORS on KSS was used, based on earlier work (Ahlström, Nyström et al. 2012) and the definitions of ORS and KSS. The following corresponding levels were used in the present study: ORS 0:  $1 \leq \text{KSS} \leq 5$ , ORS 2:  $6 \leq \text{KSS} \leq 7$ , ORS 3:  $8 \leq \text{KSS} \leq 9$ .

The video ORS scale used in the study contained three levels:

- **ORS 0: Alert**
  - Blink: normal
  - Yawn: no
  - Body position: sitting still
  - Body movements: normal
- **ORS 1: Some signs of sleepiness**
  - Blink: sporadic periods of long eyelid closure (followed by increased level of blink frequency)
  - Yawn: some
  - Body position: some situations with changing position – e.g. stretch
  - Body movements: some – arms, legs, scratch touching eyes or own touch
- **ORS 2: Severe sleepiness**
  - Blink: half-closed eyes, empty gaze
  - Yawn: yes
  - Body position: yes - change often, stretch, slumped, hanging
  - Body movements: yes - e.g. head nodding

In detail, the procedure was the following: At arrival, the participants were served refreshments and were given a short introduction to the task ahead of them and were also given a written description of the method and the procedure (see Appendix A).

Then they were shown a sequence of 10 video clips where the *true* sleepiness value, i.e. the KSS estimations by the drivers (given on the three graded ORS scale described above), was shown after each video clip in order for the raters to calibrate (practice) their estimations using ORS. Next the raters were allowed to ask questions. Following that, the first session of 18 clips started. In between each clip there was a 10 s segment where the raters were reminded to estimate the sleepiness level as well as to give their confidence in their estimations. Between the three sessions of 18 clips, the raters had a break for 10 minutes and were served refreshments.

The video clips retained from the SleepEYE I field study to be used in this study on video ORS were balanced, according to day and night and the three ORS levels based on the KSS values (see below). In order to control for learning effects of the raters, the video clips were then placed in the three different sequences such that each sequence contained an equal distribution of clips for ORS 0 to 2 and of day and night clips. Thus, there were three video clips for each of the six possible combinations of ORS level and time of day in each session (sequence). The video clips were also balanced on drivers to the best possible extent, given the fact that only some drivers estimated a more severe level of sleepiness.

The raters were recruited from two different age populations, one young adult and one middle-aged. They were on average 39 years old, the average age of the young adults being 23 years and the average of the middle-aged group being 58 years. All of the 40 raters except 6 (of which 5 belonged to the young group) held driving licences and had done so (on average) for 5 years in the young group and for 38 years in the middle-aged group. There were an equal number of females and males in both groups. The raters also had to answer a question if they had fallen asleep or felt very sleepy while driving during the last year. Ten of the 40 participants answered yes to this question.

Finally, the raters were informed that the video clips had been randomly selected and that there were sleepy drivers in the daytime sessions and alert drivers in the night-time sessions, but that the order and distribution were randomized (note that only the order was randomized, but the distribution was not, as described above).

### 5.1.1 Video data

The videos used in this study were, generally, of poor quality. However, they are representative of the video quality in the euroFOT data. Figure 20 shows one example from night-time driving (left) and one example from daytime driving (right). An apparent problem in the night-time driving frame is too much light falling on the driver's face. This was a problem throughout the video data in this study.



Figure 20: Examples of the video clips. Left: night-time driving, right: daytime driving. Videos from the SleepEYE I field experiment.

## 5.2 Results

The average percentage of correctly classified video clips was 39% with a standard error of the mean of 1.0%. Given that the ORS scale consisted of three levels, 33.3% correct classification corresponds to chance (i.e. someone picking ORS category at random for each video clip).

The relative frequency of correctly classified percentage of clips per rater can be seen in Figure 21.

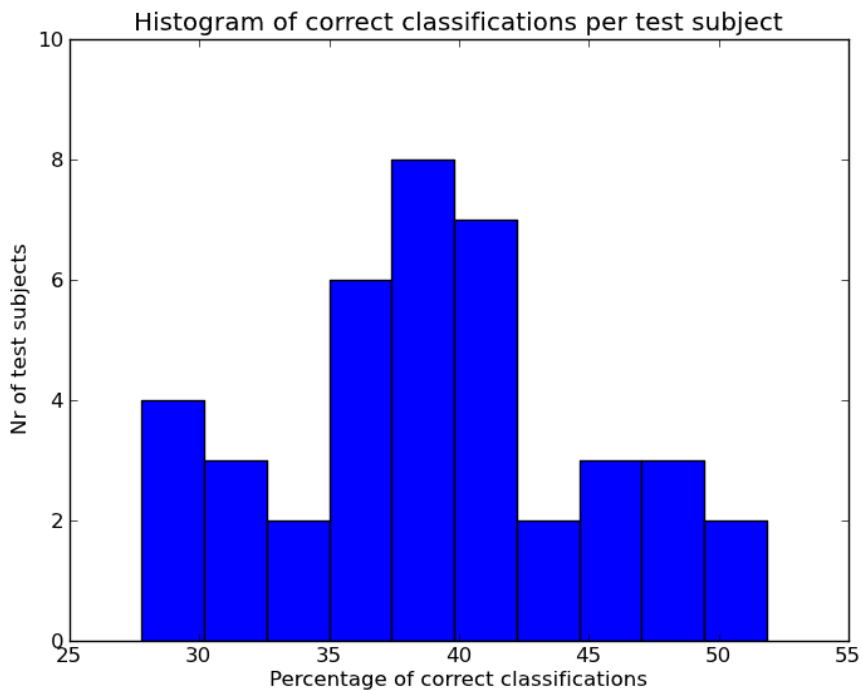


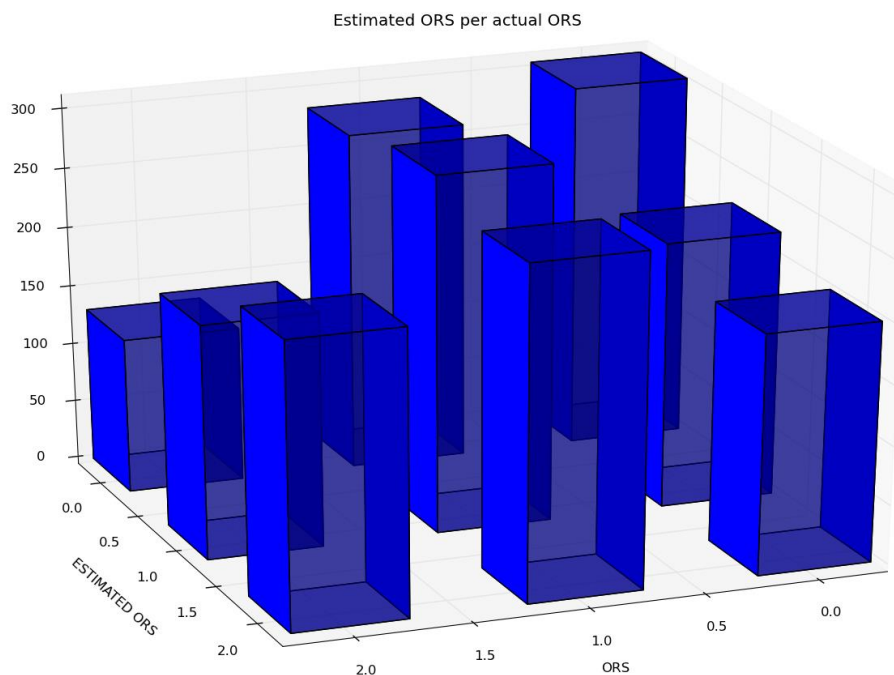
Figure 21: Histogram of the raters' percentage of correct classifications.

The average correct percentage of classifications by the raters for the three ORS levels and time of day (night or day) are shown in Table 6.

*Table 6: The average percentage of correct classifications by the raters for the six conditions.*

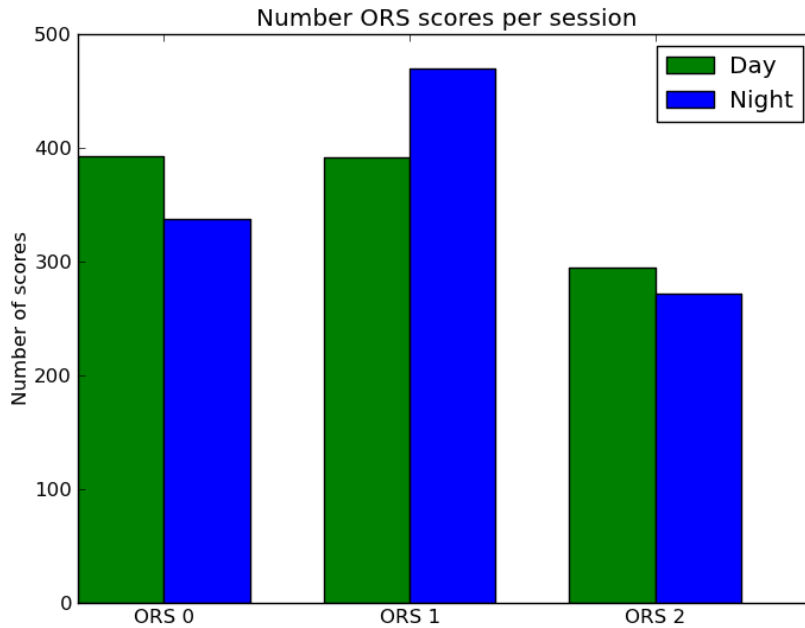
|                | <b>ORS 0</b>        | <b>ORS 1</b>        | <b>ORS 2</b>        | <b>AVERAGE</b>      |
|----------------|---------------------|---------------------|---------------------|---------------------|
| <b>DAY</b>     | 44.2% ( $\pm 2.1$ ) | 38.1% ( $\pm 2.6$ ) | 33.6% ( $\pm 2.7$ ) | 38.6% ( $\pm 1.6$ ) |
| <b>NIGHT</b>   | 40.6% ( $\pm 2.5$ ) | 45.0% ( $\pm 2.5$ ) | 32.8% ( $\pm 2.3$ ) | 39.4% ( $\pm 1.1$ ) |
| <b>AVERAGE</b> | 42.4% ( $\pm 1.8$ ) | 41.5% ( $\pm 2.1$ ) | 33.2% ( $\pm 2.0$ ) | 39.0% ( $\pm 1.0$ ) |

It is evident from Table 6 that ORS 0 was easier to score than ORS 2. As also can be seen there was no difference in rating scores between the day and the night clips. In Figure 22 the distribution of number of ORS estimations (estimated ORS by the raters) per true ORS level (i.e. formed by considering the drivers' subjective level of sleepiness) is shown.



*Figure 22: The distribution of the number of ORS ratings per true ORS level.*

The distribution of the ORS scores by the raters per time of day (i.e. day or night driving) is shown in Figure 23.



*Figure 23: Distribution of the ORS ratings by the raters per time of day*

The differences in Table 6 above between the ORS values estimated for the two different sessions (i.e. daytime driving and night-time driving) are small but there are some trends: A higher number of video clips from daytime driving was scored as being alert driving compared to a higher number of night-time driving video clips being scored as sleepy driving. Of the three ORS levels, ORS 2 was the least scored and ORS 1 the most scored.

The percentage of correct classifications by the young and the middle-aged raters was 40% (SEM 1.6%) and 38% (SEM 1.2%), respectively. There was no substantial difference in scoring performance between the two age groups.

As described above, for each ORS estimation made the raters gave a confidence value indicating how sure (or unsure) they were of that particular estimation. In Figure 24 the percentage of correctly classified video clips are plotted against confidence value.



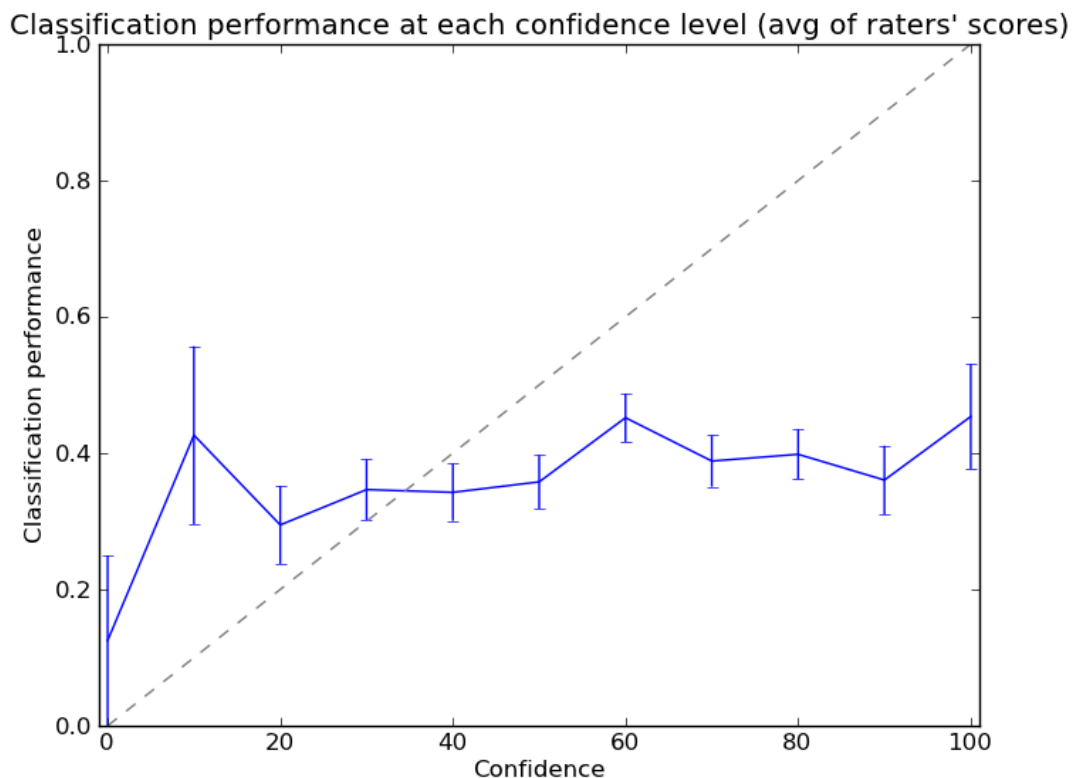


Figure 24: The percentage of correctly classified video clips (classification performance) in relation to the confidence values given by the raters.

The dashed line represents the expected classification performance per confidence level if the raters had given perfect confidence estimations. However, the actual performance at each confidence level (blue line) clearly shows that the only real difference between the confidence levels are between confidence 0 and confidence higher than zero.

Finally, in a similar approach to the attempt by Wierwille and Ellsworth (Wierwille & Ellsworth 1994), to quantify inter-individual differences between the raters, the covariance matrix of all raters' estimations of the 54 video clips was computed. The average for the Pearson's  $r$  correlations between all raters (i.e. the average of all the values above the main diagonal in the covariance matrix) was then formed and found to be 0.22. The correlation coefficients between raters varied quite a lot, in fact, the standard deviation was found to be 0.15. (The correlation coefficients were made additive by way of Fishers'  $z$  transformation, something Wierwille and Ellsworth seemingly failed to consider in their work.)

### 5.3 Discussion

The results presented above show that, under the circumstances of the present study, estimating a driver's level of sleepiness by watching a 1-minute video recording of the driver is, in fact, quite difficult. Of course, the quality of the videos used may be an important limitation to the results.

Studying the distribution of estimated ORS per actual ORS level it can be seen that for actual ORS 0 the raters' estimations of ORS were correct for 42% of the video clips, but completely wrong (i.e. estimated ORS 2 for actual ORS 0) in only 18% of the cases.

The situation was similar for actual ORS 2. For actual ORS 1, the distribution of estimated ORS levels was almost uniform over the three levels.

Regarding the confidence estimates there is a weak trend (especially if the performance at confidence level 10% is disregarded as an outlier) that the higher the confidence the higher the probability of a correct ORS estimation. However, there is very much uncertainty and this is probably what explains the outlier at confidence 10%. Still, the ratings for the daytime videos, where the quality of the videos often was rather good, were not much better than the ratings for the night-time videos.

By studying the actual ORS estimations grouped by time of day of the video clips, it was observed that the raters gave higher ORS estimations (ORS 1 or 2) for the night-time videos and lower ORS estimations (ORS 0) for the daytime videos albeit the difference was very small, almost negligible.

An interesting question concerns the learning aspect and whether it is possible to train the raters in order for them to improve their estimates. In the work by Wierwille and Ellsworth (Wierwille & Ellsworth 1994) the 6 raters they studied were claimed to be “trained raters”, simply by being college students at behavioural science programs. The raters who participated in the study presented here were recruited from the ordinary population with no regard to education. The students in Wierwille’s and Ellsworth’s study did not, however, receive any training of estimating driver sleepiness whereas the participants in the study reported here actually had one training session that was not included in the analyses detailed above.

Two important aspects of the present study concern: (1) the video quality and (2) the usage of KSS to represent the true ORS value. Regarding the first aspect, other camera systems may provide much better videos of the driver’s face and therefore make the ORS estimations much more reliable. However, in this work the question was whether the videos recorded in the euroFOT database could be used to score ORS. Therefore, videos recorded with a euroFOT vehicle were used in this study. It is of course very interesting to work with higher quality video in order to try to answer the question on whether video ORS is at all feasible. Moving on to the topic of true ORS values it should, of course, not be expected that the subjective estimations (i.e. KSS estimations which were used to define the true ORS values for each video clip) perfectly reflect what may be observable for a rater scoring ORS. In other words, there may be differences in subjectively experienced levels of sleepiness and in behavioural signs of sleepiness that can be observed by others. It may be argued that in order to use an approach such as ORS, to study driver sleepiness related questions, more fundamental understanding of how behavioural signs of sleepiness relate to physiological signs of sleepiness and subjective experiences of sleepiness is needed.

Finally, the length of the video sequences may have had an impact on the results. In particular, behavioural signs of sleepiness such as yawning and nodding are, perhaps, infrequent events. Ratings of sleepiness based on such events may be made more reliable with longer video clips. However, lengthening the video clips carries an increased cost of scoring video data. Still, a study involving longer video clips would be interesting future work.

## 6 Analysis of driver sleepiness in euroFOT data

The original intention was to apply the suggested sleepiness indicators as well as the video ORS to the euroFOT data, and to analyse and compare the outcomes from using these two methods. However, since the results of the video ORS study were not satisfying the analysis was revised to include only the sleepiness indicators.

### 6.1 Method

#### 6.1.1 Selection of data from the euroFOT database

The same inclusion criteria as were used when evaluating the blink detection algorithm in Section 3.3.2 were used for the final analysis of driver sleepiness in euroFOT data. In other words, only trips that fulfilled the following criteria were included:

- Eye tracking data collected with the last version of the eye tracker software
- Motorway driving
- Blink frequency in the range of 5 - 75 blinks/min

In the euroFOT database 180 trips fulfilled the selection criteria (see also section 3.3.2). In total, these trips corresponded to 116 hours of data and they were driven by 20 drivers.

#### 6.1.2 Application of sleepiness indicators on euroFOT data

As described in section 4.2.1, the sleepiness indicators applied to the euroFOT data were:

- Mean blink duration (camera-based)
- Number of line crossings

The definition (i.e. the detection algorithm) of mean blink duration used was identical to that applied to SleepEYE I data. The definition of the number of line crossings was however somewhat different. In SleepEYE I data potential line crossings were detected by an algorithm and the results were manually checked by watching the video films and the lane tracker data, in order to remove false positives (which may be present e.g. because of poor lane markings or other factors that are hard to predict). The euroFOT dataset was too large to use the same approach. Therefore, line crossings were obtained from the lane departure warning (LDW) system instead. A line crossing was thus defined to occur when an LDW warning was issued without a voluntary lane change taking place. Voluntary lane changes were identified by the event marking "lane change" that was available in the euroFOT database.

## 6.2 Results

Figure 25 shows a scatter plot of the number of line crossings and mean blink duration. The blue line corresponds to the simple classification function suggested in section 4.2.2. It can be seen that 4 of 2320 data points are outside the line. These data points (that are outside the line) are coming from one driver.

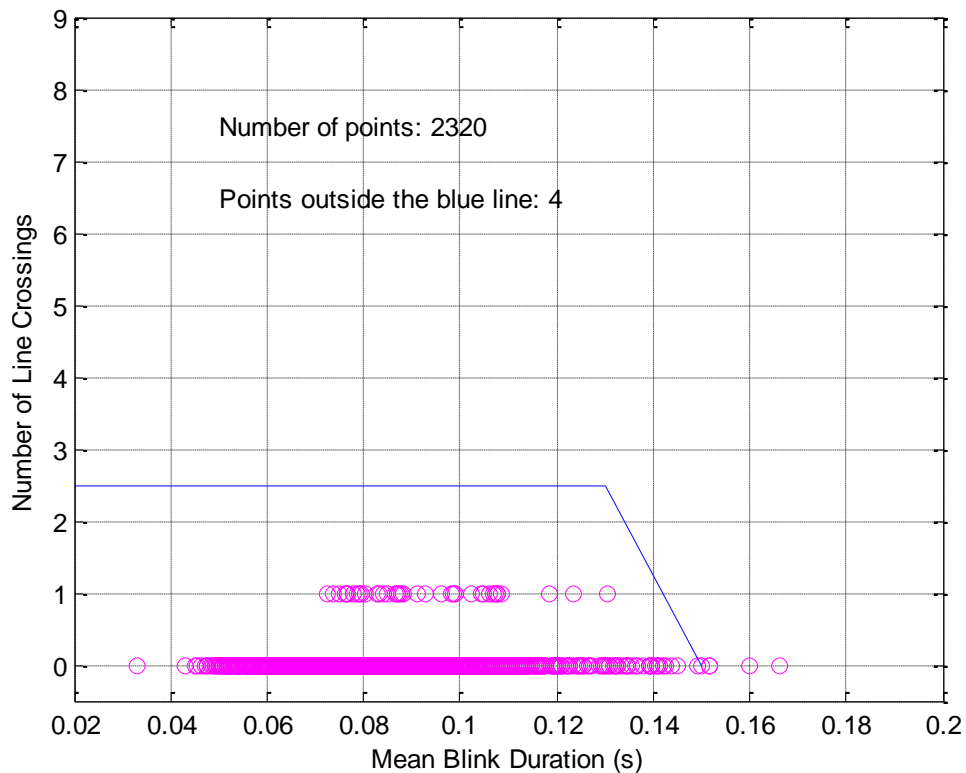


Figure 25: Scatter plot of number of line crossings (per 5 min) and mean blink duration (s) in the euroFOT dataset. The blue line indicates an example of a classification function.

Two of the data points (outside the line) come from one driver in one trip. The driver is dialing on a cell phone and looking away from the road for a relatively long time, and then continues to talk on the phone. The trip is 20 minutes long, and these blinks are occurring 4 (corresponds to 17:41) and 5 minutes (corresponds to 17:42) after the trip start (17:37).

The third data point comes from the same driver but in a different trip at 16:22 (18 minutes after the trip start, the trip is 30 minutes long). In this trip the driver is talking on the cell phone constantly and driving with one hand.

The fourth data point also comes from the same driver but in yet another trip at 19:35 (25 minutes after the trip start, the trip is 40 minutes long). The driver is talking to the passenger next to him.

### 6.3 Discussion

The results from the analysis of euroFOT data are meagre – less than one per cent of the data was above the suggested sleepiness threshold. From the qualitative analysis it can be concluded that these data points most likely are false positives. It was found that the driver was speaking in these situations, and this is a difference compared to the SleepEYE I experiments, where the driver was not allowed to speak (a detailed description of the experiment can be found in (Fors, Ahlström et al. 2011)). Thus, this kind of false positives could not be identified or taken into account when suggesting indicators based on the SleepEYE I data.

Another difference is the fact that line crossings were obtained from the LDW system in the euroFOT data. This means that a warning was issued every time a line crossing occurred, which may have had an alerting effect on the driver (although only a short-time effect can be expected) or may have caused the driver to stop driving.

The most important difference between euroFOT data and SleepEYE I data is probably that the drivers are expected to be much sleepier in SleepEYE. Figure 26 shows a scatter plot of mean blink duration and time of day for the 180 euroFOT trips. The vast majority of the trips were undertaken during daytime, when the sleep pressure is expected to be low. In contrast, half of the driving sessions in SleepEYE I was conducted between midnight and 5 o'clock in the morning.

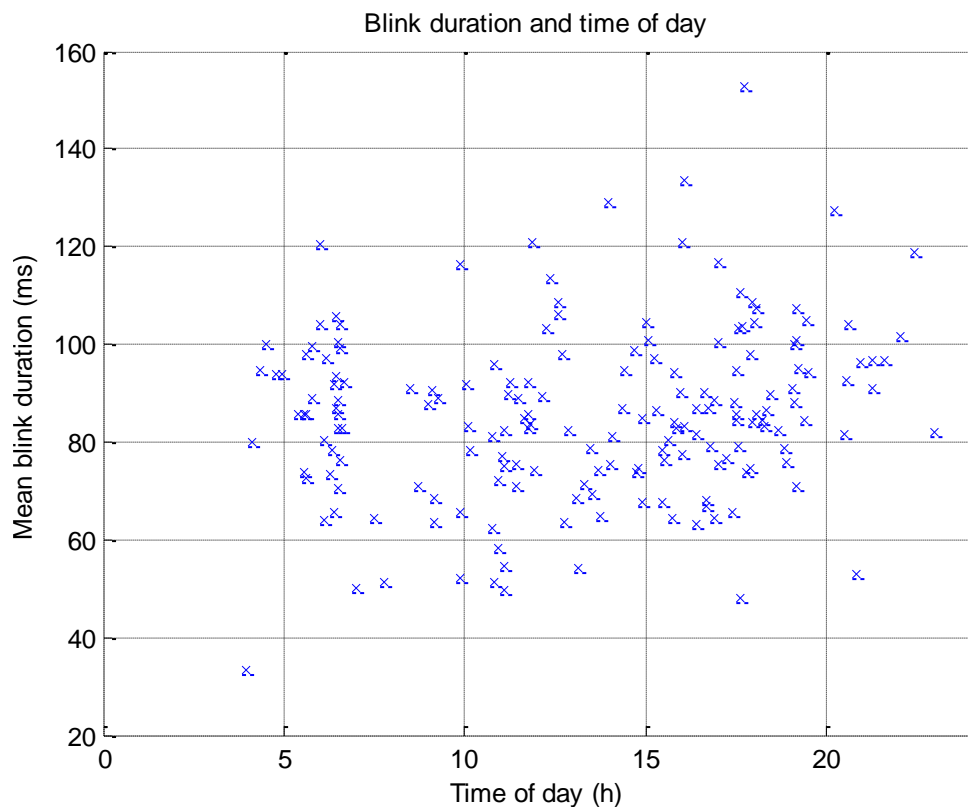
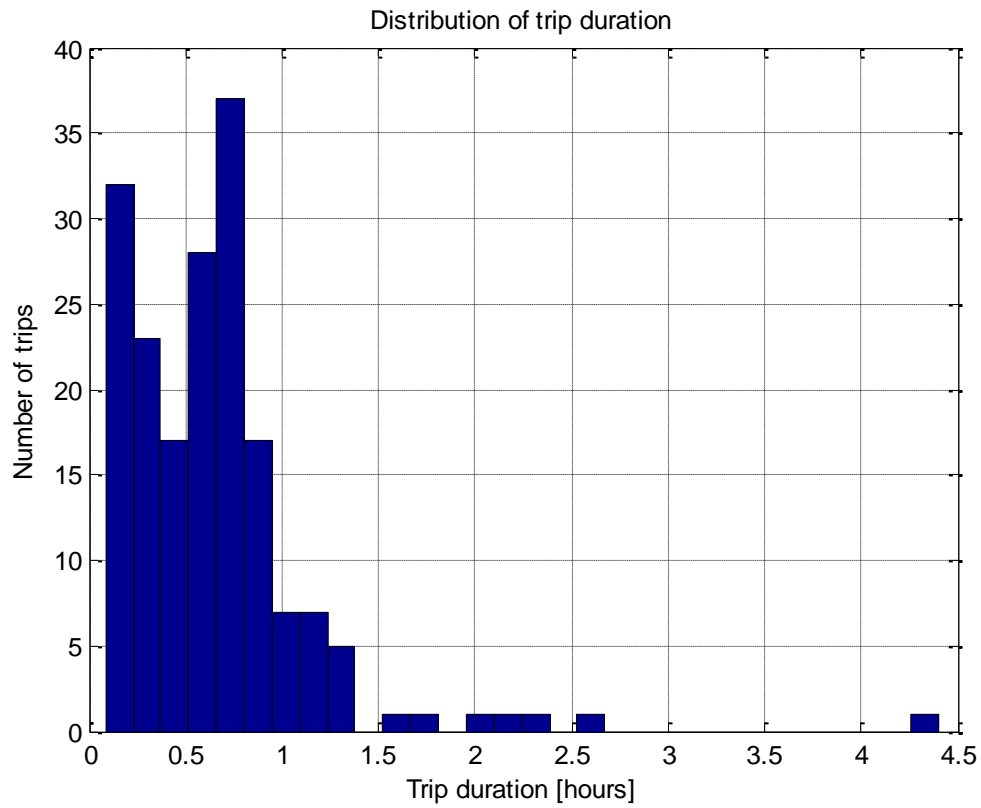


Figure 26: Mean blink duration and time of day for the 180 trips selected for the analysis from the euroFOT database.

Furthermore, sleepiness usually increases with trip duration, and there were in general shorter trips in euroFOT than in SleepEYE I. Figure 27 shows the distribution of trip durations for euroFOT, where it can be seen that most trips were shorter than 1 h. In SleepEYE I the trip durations were about 1.5 h. Another difference is that the participants in SleepEYE were not allowed to drink any caffeine containing beverages, nor were they allowed to listen to the radio or do anything else that would counteract their sleepiness.

Accordingly, the very low rate of data points above the sleepiness level might not be an unexpected result. On the other hand, the number of false negatives cannot be assessed. Since only about 30% of the sleepy data points in SleepEYE I was correctly classified a large amount of the sleepy data can be expected to overlap with the alert data. Thus, there may be sleepy data in euroFOT that could not be identified. There is however no

way of solving this when only the two indicators blink duration and line crossings are used. Modifying the thresholds is probably not a good idea either, since all data is concentrated to a rather small area in Figure 25.



*Figure 27: Distribution of trip duration for the 180 trips selected for the analysis from the euroFOT database.*

## 7 Summary and recommendations

The main results of work packages 2 and 3 of the SleepEYE II project (as described in detail in earlier sections of this report) are as follows:

(1) The software for detecting blinks in the eye tracking camera system has been improved compared to the system used in the SleepEYE I project. The problem of the large number of blinks that the previous system failed to identify has been mitigated to some extent, along with the problem of downward gazes that were often incorrectly identified as long blinks. After several iterative steps of improving the detection algorithm the software ended up being completely reworked and a new detection algorithm has thus been devised and implemented within the framework of SleepEYE II. The new detection algorithm had an acceptable detection rate when applied to data from the SleepEYE I field test, but for euroFOT data the number of identified blinks was unreasonably low in about half of the trips. There is thus a need for further improvements of the blink detection algorithm.

(2) An in-depth study on indicators of driver sleepiness was carried out, using data collected in a controlled field experiment (SleepEYE I) on driver sleepiness, with the purpose of employing the indicators to study driver sleepiness in the euroFOT database. However, of the most promising indicators, namely  $KSS_{estSWP}$ , number of line crossings, and EOG-based measures, only line crossings could be considered when analysing the euroFOT data. Reasons are that in euroFOT the information required by SWP (i.e. time of day at the time of driving as well as sleep history) was not collected and, of course, EOG was not recorded. The EOG-based measures were replaced by the camera-based counterparts and a classifier was then devised, based on these indicators (number of line crossings and camera-based mean blink duration), and implemented to detect sleepy driving in the euroFOT data. When applied to SleepEYE I data the classifier was found to have good specificity while the sensitivity of the classifier was not so good. For euroFOT data, no true data on the drivers' sleepiness level were available and it was therefore not possible to evaluate the performance of the classifier, but an explorative analysis showed that only very few data points were classified as coming from occurrences of sleepy driving. This may be reasonable since most trips were conducted during daytime, but it is a somewhat disappointing result for the project.

(3) A study was carried out on using video data of the drivers in order to estimate the drivers' level of sleepiness. Forty participants (observers) rated 54 one-minute video clips of an equal number of sleepy and alert drivers on a scale with three levels (alert, first signs of sleepiness, very sleepy). The study showed that performing such observer rated sleepiness (ORS) estimations on drivers is extremely difficult. The videos available in euroFOT are of rather poor quality which, likely, significantly limits the possibility of making reliable observer rated sleepiness estimations.

It can be concluded that studying driver sleepiness in FOT data is difficult, for several reasons:

- Eye based indicators suffer from detection errors and low detection rate (e.g. because of sunlight, downward gazes, and head movements).
- Driving based indicators are influenced by e.g. road curvature, traffic density and the presence of driver support systems (such as lane departure warnings).
- Models of sleepiness cannot be used since no information on hours slept and time awake is available.

- Video scoring is not reliable, at least not given the quality of the available video recordings.

## 7.1 Lessons learned

The quality of data, mainly with respect to eye tracking (blinks) and video recordings, turned out to be a considerable issue. A lot of time was spent improving and evaluating the quality of blink data and, as a consequence, less time was left to the actual aims of the study, i.e. developing methods for identification of sleepy driving in FOT data.

The poor quality of the video recordings may explain the poor results of the video ORS method. The lack of a reliable video ORS method clearly limited the possibilities to analyse driver sleepiness in euroFOT data, since it was expected that the videos probably would be the most reliable source of information with regard to the drivers' sleepiness level.

The lesson that can be learnt from the project is that the quality of data is very central and very important. It is difficult to give any general recommendations on how to avoid data quality problems, but it is probably a good idea to always allocate time for quality checks and quality issue handling.

## 7.2 Recommendations for future FOT studies

Based on the findings in the work carried out within the SleepEYE II project, it may be suggested that in order to make studying of driver sleepiness possible in FOT studies the following should be taken into consideration:

- The Sleep/Wake Predictor (SWP) model has proven to be a very good estimator of sleepiness levels. Therefore, it is highly recommended that information about the drivers' sleep is collected when possible (possibly by using sleep diaries or technology such as actigraphs).
- Improved video quality is needed in order to not limit the observer rated estimations of sleepiness which are carried out using the video recordings of the drivers' faces.
- Further development of eye tracking technology, especially with respect to robustness, may be needed in order to make reliable analyses of blink behaviour based indicators of sleepiness.
- Sleepiness indicators that are used must be robust with respect to the naturalistic setting, i.e. varying driving environments and weather conditions, natural driver behaviours (e.g. speaking, listening to the radio, etc.) and individual driving styles.
- In order to be able to study driver sleepiness in future FOTs, the sleepiness perspective should be included already when planning the study, for two reasons. First, trips where the probability of sleepiness is high must be included, e.g. long trips and/or trips late at night. Second, some information about the drivers' sleep and sleepiness, such as sleep diaries or KSS, should be collected during the test period.



### 7.3 Contribution to ViP

This project has contributed to the knowledge base within ViP by:

- Linking a simulator experiment to a controlled real road study which, in turn, is linked to a FOT study.
- Improving the embedded camera system and thus providing a tool for unobtrusive and low-cost measurements of blink behaviour.
- Evaluating a method for observer rated sleepiness that may be employed, given further research, as a reference for sleepiness in the development of driver sleepiness detection and warning systems.

## References

- Ahlström, C., Nyström, M., Holmqvist, K., Fors, C., Sandberg, D., Anund, A., Kecklund, G. & Åkerstedt, T. (2012). *Fit-for-duty test for estimation of drivers' sleepiness level: Eye movements improve the Sleep/Wake predictor* submitted.
- Akerstedt, T. (2000). *Consensus statement: fatigue and accidents in transport operations*. *J Sleep Res* **9**(4): 395.
- Anund, A., Fors, C., Hallvig, D., Åkerstedt, T. & Kecklund, G. (2013). *Observer Rated Sleepiness and Real Road Driving: An Explorative Study*. *PLoS ONE* **8**(5).
- Anund, A., Kecklund, G., Vadeby, A., Hjälmdahl, M. & Åkerstedt, T. (2008). *The alerting effects of hitting a rumble strip - a simulator study*. *Accid Anal Prev* **40**: 1970-1976.
- Barr, L.C., Yang, C.Y.D., Hanowski, R.J. & Olson, R. (2005). *Assessment of driver fatigue, distraction, and performance in a natural setting*. *Transportation Research Record* **1937**: 51-60.
- Davenne, D., Lericollais, R., Sagaspe, P., Taillard, J., Gauthier, A., Espie, S. & Philip, P. (2012). *Reliability of simulator driving tool for evaluation of sleepiness, fatigue and driving performance*. *Accid Anal Prev* **45**: 677-682.
- Dingus, T.A., Neale, V.L., Klauer, S.G., Petersen, A.D. & Carroll, R.J. (2006). *The development of a naturalistic data collection system to perform critical incident analysis: An investigation of safety and fatigue issues in long-haul trucking*. *Accid Anal Prev* **38**: 1127-1136.
- Fors, C., Ahlström, C., Sörner, P., Kovaceva, J., Hasselberg, E., Krantz, M., Grönvall, J.-F., Kircher, K. & Anund, A. (2011). *Camera based sleepiness detection – Final report of the project Sleep EYE*. Linköping, Sweden: Swedish National Road and Transport Research Institute (ViP Publication 2011-6).
- Gillberg, M., Kecklund, G. & Åkerstedt, T. (1996). *Sleepiness and performance of professional drivers in a truck simulator - comparisons between day and night driving*. *J Sleep Res* **5**(1): 12-15.
- Horne, J.A. & Reyner, L.A. (1995). *Sleep related vehicle accidents*. *BMJ* **310**(6979): 565-567.
- Häkkinen, H., Summala, H., Partinen, M., Tiihonen, M. & Silvo, J. (1999). *Blink duration as an indicator of driver sleepiness in professional bus drivers*. *Sleep* **22**(6): 798-802.
- Ingre, M., Åkerstedt, T., Peters, B., Anund, A., Kecklund, G. & Pickles, A. (2006). *Subjective sleepiness and accident risk avoiding the ecological fallacy*. *J Sleep Res* **15**: 142-148.
- Jammes, B., Sharabaty, H. & Esteve, D. (2008). *Automatic EOG analysis: A first step toward automatic drowsiness scoring during wake-sleep transitions*. *Somnologie* **12**: 227-232.
- Kaida, K., Takahashi, M., Akerstedt, T., Nakata, A., Otsuka, Y., Haratani, T. & Fukasawa, K. (2006). *Validation of the Karolinska sleepiness scale against performance and EEG variables*. *Clin Neurophysiol* **117**(7): 1574-1581.

- Kecklund, G. & Åkerstedt, T. (1993). *Sleepiness in long distance truck driving: an ambulatory EEG study of night driving*. *Ergonomics* **36**(9): 1007-1017.
- Liu, C.C., Hosking, S.G. & Lenné, M.G. (2009). *Predicting driver drowsiness using vehicle measures: Recent insights and future challenges*. *Journal of Safety Research* **40**(4): 239-245.
- Lowden, A., Anund, A., Kecklund, G., Peters, B. & Åkerstedt, T. (2009). *Wakefulness in young and elderly subjects driving at night in a car simulator*. *Accid Anal Prev* **41**: 1001-1007.
- Maycock, G. (1997). *Sleepiness and driving: the experience of U.K. car drivers*. *Accid Anal Prev* **29**(4): 453-462.
- NTSB (1999). *Evaluation of the U.S. department of transportation efforts in the 1990s to address operator fatigue*. Washington D.C., U.S. (Safety Report)
- Philip, P., Sagaspe, P., Moore, N., Taillard, J., Charles, A., Guilleminault, C. & Bioulac, B. (2005a). *Fatigue, sleep restriction and driving performance*. *Accid Anal Prev* **37**(3): 473-478.
- Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., Charles, A. & Bioulac, B. (2005b). *Fatigue, sleepiness, and performance in simulated versus real driving conditions*. *Sleep* **28**(12): 1511-1516.
- Philip, P., Vervialle, F., Le Breton, P., Taillard, J. & Horne, J.A. (2001). *Fatigue, alcohol, and serious road crashes in France: factorial study of national data*. *BMJ* **322**(7290): 829-830.
- Radun, I. & Summala, H. (2004). *Sleep-related fatal vehicle accidents: characteristics of decisions made by multidisciplinary investigation teams*. *Sleep* **27**(2): 224-227.
- Reyner, L.A. & Horne, J.A. (1998). *Falling asleep whilst driving: are drivers aware of prior sleepiness?* *Int J Legal Med* **111**(3): 120-123.
- Sagaspe, P., Taillard, J., Åkerstedt, T., Bayon, V., Espie, S., Chaumet, G., Bioulac, B. & Philip, P. (2008). *Extended driving impairs nocturnal driving performances*. *PLoS ONE* **3**(10).
- Sandberg, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J.G., Wahde, M. & Åkerstedt, T. (2011a). *The characteristics of sleepiness during real driving at night - a study of driving performance, physiology and subjective experience*. *Sleep* **34**(10): 1317-1325.
- Sandberg, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J.G., Wahde, M. & Åkerstedt, T. (2011b). *Sleepy driving on real roads and in a simulator - an absolute comparison*. Submitted (available in Sandberg, D., *Detecting driver sleepiness*, dissertation, Chalmers university of technology, 2011).
- Sandberg, D., Åkerstedt, T., Anund, A., Kecklund, G. & Wahde, M. (2011). *Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators*. *IEEE Transactions on Intelligent Transportation Systems* **12**(1).
- Simon, M., Schmidt, E., Kincses, W., Fritzsche, M., Bruns, A., Aufmuth, C., Bogdan, M., Rosenstiel, W. & Schrauf, M. (2011). *EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions*. *Clin Neurophysiol* **122**(6): 1168-1178.

- Tefft, B.C. (2012). *Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999–2008*. *Accid Anal Prev* **45**: 180-186.
- Wierwille, W.W. & Ellsworth, L.A. (1994). *Evaluation of driver drowsiness by trained raters*. *Accid Anal Prev* **26**(5): 571-581.
- Åkerstedt, T., Folkard, S. & Portin, C. (2004). *Predictions from the Three-Process Model of Alertness*. *Aviation, Space, and Environmental Medicine* **75**(Suppl 1, March): A75-A83.
- Åkerstedt, T. & Gillberg, M. (1990). *Subjective and objective sleepiness in the active individual*. *International Journal of Neuroscience* **52**: 29-37.
- Åkerstedt, T., Peters, B., Anund, A. & Kecklund, G. (2005). *Impaired alertness and performance driving home after the night shift: a driving simulator study*. *J Sleep Res* **14**(1): 17-20.

## Video-ORS instructions

Below are the instructions (in Swedish) given to the participants in the video-ORS study.

### Försök att skatta sömnhet hos förare på videofilm

#### Bakgrund

I dagsläget så pågår det flera stora projekt vars syfte är att bättre förstå vad som kan leda till farliga situationer vid bilkörning. Exempel på sådana faror är distraktion och sömnhet. Det är vanligt att man i dessa projekt samlar in stora mängder videofilmer från kameror placerade i försökspersoners bilar. Vi undrar nu om man tillförlitligt kan uppskatta hur sömning en förare har varit genom att titta på en sådan videoinspelning.

#### Din uppgift

Du kommer att få se totalt 54 st 1-minuters videoklipp. Dessa videoklipp är inspelade med en kamera som suttit i en försöksbil och som filmat olika förare under riktig körning. Din uppgift är att för varje videoklipp försöka, så gott du kan, uppskatta hur sömning föraren på videon är. Du gör dessa skattningar på en tregradig skala som beskrivs på ett separat papper. För varje sömningsskattning du gör så ska du också försöka uppskatta hur säker du är på att denna sömningsskattning stämmer med förarens verkliga trötthet. Denna säkerhetsuppskattning gör du på en skala från 0 till 100 % (i steg om 10 %) där 0 % motsvarar att du inte har någon aning och 100 % betyder att du helt säker på att din skattning stämmer med förarens verkliga trötthet.

Du kommer först att få se 10 st 1-minuters videoklipp som träning. För dessa videoklipp kommer förarnas egna uppskattningar om hur trötta de var att presenteras, men först får du själv bedöma vad du tror.

#### Att tänka på

När du genomför skattningarna är det viktigt att du ser varje 1-minuters klipp som ett unikt klipp. Tänk också på att göra din sömningsskattning först mot slutet av varje klipp.

Ordningen på klippen och hur sömniga personerna är presenteras slumpmässigt.

Notera att videofilmerna som du kommer att få se har spelats in både i dagsljus och i mörker, men att förarna kan vara trötta och pigga både dag som natt.

Ljusförhållandena (och elektroderna) kan ibland göra att det är svårt att alls se föraren, men gör ändå ditt bästa utifrån den information som finns i videon att bedöma hur trött föraren är. Det finns även annan utrustning i bilen som gör ljuset kan blinka i vissa klipp.

Ni får gärna prata med varandra i pauserna, men undvik att diskutera själva försöket utan vänta med detta till efter sista sessionen.

***Tack för att du är med och hjälper oss!***

### Skattningsskala sömnighet

- Välj den nivå som bäst stämmer överens med förarens beteende
- Alla kriterier behöver inte vara uppfyllda för att välja en viss nivå

| 0: Pigg  | 1: Sömnig, men ej ansträngande att vara vaken  | 2: Mycket sömnig, ansträngande att vara vaken  |
|--|--|--|
| <p><b>Blinkningar:</b> Normala</p> <p><b>Gäspningar:</b> Inga</p> <p><b>Kropposition:</b> Sitter stilla</p> <p><b>Kropps rörelser:</b> Normala</p> | <p><b>Blinkningar:</b> Enstaka perioder med halvslutna ögon (som efterföljs av snabba blinkningar)</p> <p><b>Gäspningar:</b> Några få eller inga</p> <p><b>Kropposition:</b> Enstaka tillfällen då man byter kropposition – t ex att man sträcker sig</p> <p><b>Kropps rörelser:</b> Enstaka tillfällen - armar, ben, kliningar, gnuggningar i ögon, egen beröring</p> | <p><b>Blinkningar:</b> Halvslutna ögon, tom blick</p> <p><b>Gäspningar:</b> Ja</p> <p><b>Kropposition:</b> Byter ofta -sträcker sig, hopsjunken, hängande</p> <p><b>Kropps rörelser:</b> Ja – tillnickningar, "tappar huvudet"</p> |





# ViP

## Virtual Prototyping and Assessment by Simulation

ViP is a joint initiative for development and application of driving simulator methodology with a focus on the interaction between humans and technology (driver and vehicle and/or traffic environment). ViP aims at unifying the extended but distributed Swedish competence in the field of transport related real-time simulation by building and using a common simulator platform for extended co-operation, competence development and knowledge transfer. Thereby strengthen Swedish competitiveness and support prospective and efficient (costs, lead times) innovation and product development by enabling to explore and assess future vehicle and infrastructure solutions already today.



Virtual Prototyping and  
Assessment by Simulation

Centre of Excellence at VTI funded by Vinnova and ViP partners

VTI, Scania, Volvo Trucks, Volvo Cars, Swedish Transport Administration,  
Dynagraph, Empir, HiQ, SmartEye, Swedish Road Marking Association

[www.vipsimulation.se](http://www.vipsimulation.se)

Olaus Magnus väg 35, SE-581 95 Linköping, Sweden – Phone +46 13 204000