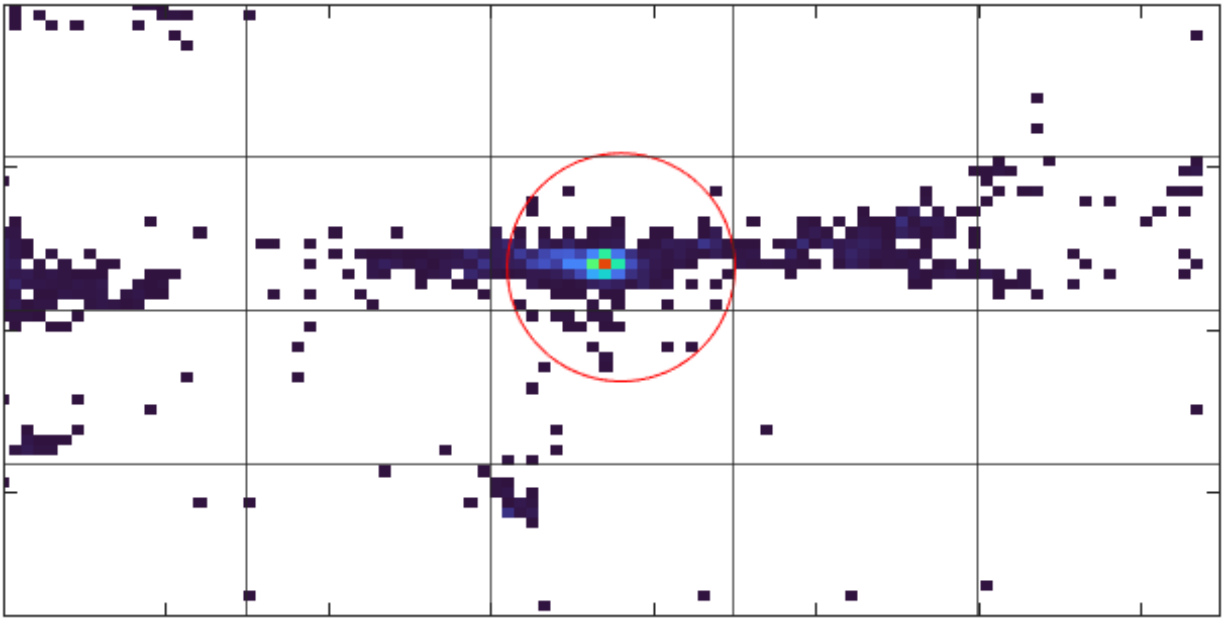
CHALMERS
UNIVERSITY OF TECHNOLOGY



Measuring drivers' gaze concentration

Comparison of metrics and eye tracking systems with respect to gaze concentration

Master's thesis in Mobility Engineering

EMMA DJERF
ELLEN RYCKENBERG

DEPARTMENT OF MECHANICAL AND MARITIME SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se

MASTER'S THESIS IN MOBILITY ENGINEERING

Measuring drivers' gaze concentration

Comparison of metrics and eye tracking systems with respect to gaze concentration

EMMA DJERF
ELLEN RYCKENBERG

Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2023

Measuring drivers' gaze concentration
Comparison of metrics and eye tracking systems with respect to gaze concentration
EMMA DJERF, ELLEN RYCKENBERG

Master's Thesis 2023
Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: + 46 (0)31-772 1000

Cover:
Density plot showing data of the dispersion of gaze, collected using a wearable eye tracker system. The data was collected in the test study conducted in this thesis.

Department of Mechanics and Maritime Sciences
Göteborg, Sweden, 2023

Measuring driver's gaze concentration
Comparison of metrics and eye tracking systems with respect to gaze concentration
Master's thesis in Mobility Engineering
EMMA DJERF, ELLEN RYCKENBERG
Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Chalmers University of Technology

Abstract

This thesis aims to compare gaze concentration metrics from the literature using a remote and a wearable eye tracking system. Drivers' gaze direction has been seen to narrow, also referred to as gaze concentration, for different driver states. These driver states could either be temporary such as cognitive distraction or lasting longer such as intoxication by drugs or alcohol. Gaze concentration can thus be used as an indication of these states. A small data collection was performed, including 5 test participants using both eye tracking systems at the same time. The participants were subjected to cognitive task during the drive to experience cognitive load. The gaze concentration metrics percent road center (PRC), percent area of interest (PAI), stationary gaze entropy (SGE), transitional gaze entropy (TGE), standard deviation of radial gaze (SDRG), standard deviation of horizontal gaze (SDHG), and standard deviation of vertical gaze (SDVG) were selected for analysis based on a literature review. The metrics were analyzed both for longer and shorter time windows. Results showed that SGE and TGE, which both uses area division of the gaze dispersion, were highly sensitive measures to the size and number of areas chosen. TGE also showed a large variation across participants and between the two systems, suggesting that it is not a robust metric. SDVG was shown to be less sensitive than the other metrics, suggesting that this metric was not suitable to capture gaze concentration during cognitive load. PRC, PAI, SGE, SDRG, and SDHG all showed similar trends, indicating that these metrics captured similar types of gaze behavior during cognitive load. Due to the small dataset used in this study, stating which metric was most sensitive in capturing gaze concentration was not possible. The analysis showed that there is a notable difference between the two eye tracking systems, indicating that comparing results from different studies should be done with care unless the same eye tracking system has been used.

Key words: Gaze concentration, Cognitive distraction, Eye tracking system, Metrics, Driver state

Contents

Preface

The thesis was done in collaboration with and financed by Volvo Cars Safety Centre at Volvo Cars Corporation (VCC) in Torslanda. The thesis was carried out from January to June of 2023 within the Division of Vehicle Safety at Chalmers University of Sweden. We would like to express our gratitude to Professor Marco Dozza, examiner of the thesis but most of all the key person of creating the opportunity for us to write this thesis in collaboration with the industry partner.

To Martin Adolfsson, Liangyu Wang and all test participants, you are highly appreciated for the support, equipment, and engagement, making our test study possible. Thank you to our co-workers and co-thesis students at VCC for making our workdays more enjoyable and filled with laughter.

Finally, we would like to specifically thank our two industry supervisors Emma Nilsson and Emma Tivesten, for all their help, guidance, and countless number of comments and positive feedback on our work and report writing. Without them this thesis would not have been what it is today.

Göteborg June 2023

EMMA DJERF, ELLEN RYCKENBERG

List of abbreviations

AOI	Area of interest
CSD	Center stack display
HUCS	Head unit coordinate system
IC	Information cluster
IR	Infra-red
LE	Left exterior (Defined AOI for the 2-camera system)
LSM	Left-hand side mirror (Defined AOI for the 2-camera system)
PAI	Percent area of interest
PRC	Percent road center
RE	Right exterior (Defined AOI for the 2-camera system)
RSM	Right-hand side mirror (Defined AOI for the 2-camera system)
RVM	Rear view mirror (Defined AOI for the 2-camera system)
SDG	Standard deviation of gaze
SDHG	Standard deviation of horizontal gaze
SDRG	Standard deviation of radial gaze
SDVG	Standard deviation of vertical gaze
SGE	Stationary gaze entropy
SNR	Signal-to-noise ratio
TGE	Transitional gaze entropy
VCC	Volvo Car Corporation
VOG	Video-oculography

1 Introduction

A driver's visual behavior tends to change characteristics when drivers' experience different type of distractions. An introduction to previous research and the study of eye movements during cognitive distraction is presented in this chapter. Finally, the aim and the two main research questions for this thesis are presented.

1.1 Background

Drivers are attentive and alert most of the time, and safely perform the driving task. However, drivers may occasionally be impaired due to for instance fatigue, intoxication, or distraction, leading to an increased risk of crash involvement. In the United States over 3000 people were killed and over 400 000 people were injured in vehicle crashes due to distraction in 2019 (National Center for Statistics and Analysis, 2021). Moreover, the 100-Car Naturalistic Driving Study, a large naturalistic data study performed in the USA, indicated that nearly 80% of crashes and 65% of near-crashes involved an inattentive driver. This included inattentions such as secondary task engagement, fatigue, driving-related inattention to the forward roadway, and non-specific eye glances away from the forward roadway (Neale et al., 2006). Additionally, the Swedish National Road Administration announced that in 2021, 22% of the fatal crashes in Sweden involved an intoxicated driver (Trafikverket, 2023). These numbers indicate the need to prevent crashes caused by impaired drivers.

Tracking driver's eye patterns enables the understanding of driver states and visual behavior. Since several driver states influence the driver's gaze behavior (Nilsson et al., 2020), eye tracking may be used to detect when the driver is no longer concentrated on the driving task. Therefore, in recent years in-vehicle eye tracking sensors have started to be applied in production vehicles. In 2022 Volvo Cars Corporation (VCC) launched the EX90, their first vehicle equipped with a system for driver understanding which uses eye tracking cameras, capacitive steering wheel, along with other vehicle signals to understand the driver's state (Volvo Cars, 2022).

The eye tracking system that the EX90 is equipped with is a vehicle-integrated two camera system (Volvo Cars, 2022). However, in research today, multi-camera systems are frequently used to measure gaze behavior (Abbasi et al., 2022, Pillai et al., 2022), typically varying between two and four cameras used. Other types of measurement systems such as wearable eye trackers have also been used for research studies (Nilsson et al., 2020, Niezgoda et al., 2015).

While several measurement techniques can be used to measure eye movements, the most widely used method is the optical approach. The method relies on light, typically infrared (IR), that is reflected from the eye and captured by a camera or other type of optical sensor. The changes in the reflections are analyzed to extract eye movements (Gonzalez-Sanchez et al., 2017).

An eye tracking system can capture a driver's visual behavior, such as gaze direction, saccades, number of fixations and blinks. Gaze behavior tends to have certain characteristics for specific traffic environments. On motorways and rural roads with low traffic density, drivers' gaze tends to mostly be fixated on a target point, 1-2

seconds of lead time on the road ahead, and on more future target points (so called look-ahead fixations) (Lehtonen et al., 2013). On the other hand, the gaze tends to move over a larger area when driving in intersections or urban environments where drivers need to be aware of more things such as other road users, traffic signs and other surrounding objects (Land., 2006).

Gaze behavior can change characteristics when a driver has an altered physical or mental state. Several studies (Hammel et al., 2002, Niezgoda et al., 2015, Victor et al., 2005) have studied cognitive distraction while driving, by subjecting drivers to non-driving related tasks. The results showed that the drivers' gaze tend to narrow towards a future target point while reducing their attention towards peripheral objects and events. Some argue that this visual behavior, typically referred to as gaze concentration, is a potential risk to traffic safety due to lack of situational awareness (Victor et al., 2015). Others highlight that gaze concentration leads to a driver's gaze being concentrated to the road center and that the driver therefore have a greater chance to detect threats in the forward roadway (Victor et al., 2015).

Previous research has found a link between gaze concentration and cognitive distraction in drivers (Shiferaw et al., 2019, Nilsson et al., 2020, Ahlström et al., 2009, Victor, 2005). In addition, other driver states have been studied regarding gaze concentration. During alcohol intoxication about a third of the drivers had gaze concentration while this was not present in sober baseline driving (Tivesten et al., 2023). Drivers who crashed while using highly reliable assisted automation and were faced with a conflict where they needed to interfere, showed at least one of the following three behaviors: low level of visual attention to the road, delayed visual response times, or high levels of gaze concentration (Tivesten et al., 2019).

Furthermore, previous literature have used several different metrics to measure gaze concentration as well as to investigate their correlation to different driver states (Shiferaw et al., 2019, Nilsson et al., 2020, Ahlström et al., 2009, Victor, 2005). It is however not clear how the different metrics compare in results to one another or how they compare in results between eye tracker systems. The focus of this thesis is therefore to compile the different metrics, collect gaze data by performing a small study using two different eye tracker systems, and investigate how well the metrics and eye trackers capture gaze concentration.

1.2 Purpose

The aim of this master thesis is to identify and compare gaze concentration metrics and eye tracking systems that can be used to study or detect driver states. The results will contribute to the understanding of how drivers' visual patterns can be reliably assessed to detect if drivers are not attentive or alert.

1.3 Research questions

The research questions that will be answered in this thesis are:

1. What metrics are used today and how well do they capture gaze concentration during cognitive load?
2. How large of a difference can be seen in gaze concentration metrics when using data from eye tracking glasses compared to a vehicle-integrated 2-camera system?
    - Are there similar trends in both systems?
    - Is there a difference in the amount of missing and invalid data?

2 Theory

2.1 Eye movements

When studying visual behavior, it is important to clarify some of the terms for eye movements. In this thesis the used terms are *gaze*, *glance*, *fixation*, and *saccade*, and these terms are described in relation to drivers' visual behavior.

The direction of *gaze* defines as the orientation of the eye in space (International Organisation for Standardization, 2020). The orientation of the eye refers to a specific area or point which the eyes are directed towards.

While driving, the gaze is directed to several areas of interest (AOI), which are pre-determined areas that are located within the visual field (International Organisation for Standardization, 2020). AOI can be larger areas such as the front windscreen, moveable targets such as oncoming traffic, or smaller stationary areas such as the rear-view mirror. The size of the AOI does not have an upper limit, but the lower limit needs to be larger than the resolution of the eye-measurement system being used (International Organisation for Standardization, 2020). When referring to AOI in vehicles, areas such as left- and right-hand side mirrors (LSM and RSM), rear view mirror (RVM), left and right exterior (LE and RE), center front (CF), right front (RF), information cluster (IC), center stack display (CSD) etc. are often used, see Figure 1.

Figure 1: Common AOI for a vehicle.

A *fixation* is defined as a short temporal hold of movement that maintains direction of the eyes to a particular point within an AOI while a *saccade* is a fast movement of the eyes that occurs when the gaze changes the fixation point within an AOI or between different AOI (International Organisation for Standardization, 2020).

Glances can be defined as a scientific construct that sums up fixations and saccades that creates a temporal maintaining of the gaze within an AOI (International Organisation for Standardization, 2020). A glance is a coarser unit of measure compared to a single fixation since a glance is summing up several continuous fixations that are spatially proximal in an AOI.

2.2 Eye trackers

Eye tracking measurements can be done using several techniques and measurement equipment to gather information about a driver's visual behavior. Remote and

wearable eye tracking are two main types of eye tracking systems that use different techniques. Remote tracking requires measures of head and eye movement to be done independently (Duchowski, 2007). These systems use multiple ocular features, such as the pupil and corneal reflection, to separate head movement and eye rotation. Wearable tracking requires measures of the orientation of the eye in space (Duchowski, 2007). By fixating the eye tracker to the head, these systems measure the eye's direction relative to the head.

Video-oculography (VOG) is used for many eye trackers and uses IR light to contrast the pupil in relation to the eyeball (Fredriksson & Wallin, 2020). There are two ways of measuring gaze direction with the use of VOG, bright pupil tracking and dark pupil tracking. Bright pupil tracking uses a placement of the IR-light source on the same axis as the camera. This allows most of the light to be reflected back to the camera and shows a bright pupil (Hansen & Ji, 2010). Dark pupil tracking uses an offset placement of the camera in relation to the IR light source, resulting in that no light passes through the pupil and is not reflected to the camera. The camera can then measure the position of the pupil to get an estimate of the gaze direction (Fredriksson & Wallin, 2020). Remote and wearable eye trackers using dark pupil tracking are presented in this section.

2.2.1 Multi-camera systems

A remote multi-camera system can be mounted in the vehicle where each camera module contains a camera and IR LED lights which illuminates the driver's face. The remote multi-camera system that was used in this thesis contains two camera modules and will hereby be referred to as 2-camera system.

The placement of the cameras enables multiple angles of the driver's eyes to be captured. One camera module is placed under the instrument panel behind the steering wheel, capturing the eyes of the driver from a direct forward view. The other camera module is placed on top of the dashboard in the middle of the vehicle, facing the driver.

The 2-camera system use a so called "online calibration", where each frame is calibrated in real-time. The output data is defined in the vehicle-coordinate system and the data is sampled at a frequency of 60 Hz.

2.2.2 Eye tracking glasses

The wearable eye tracking system that was used in this thesis was the Tobii Pro Glasses 3, hereby referred to as Tobii glasses. The glasses use cornel reflection and measures the orientation of the eye in space (Tobii, 2023). The set-up consists of a head unit (glasses), a recording unit, and a controller application on an external computer or smartphone.

The head unit consists of (1) two eye tracking cameras for each eye, (2) a nose pad that can be changed with additional sizing to ensure the user's eyes being positioned in the middle of the glasses, (3) a head unit cable that is connected to the recording unit, (4) a front facing HD scene camera, (5) a microphone, and (6) eight IR illuminators for each eye, see Figure 2 (Tobii AB, 2023). A head strap is attached to

the head unit to minimize movements of the unit during data collection and ensure a good fit on the user's head. The head unit samples data at a frequency of 50 Hz and is also equipped with a gyroscope and an accelerometer sampling at 100Hz.

Figure 2: The Tobii glasses head unit (Tobii AB, 2023).

Each new user and new recording start with a calibration to ensure accurate data collection. The user holds up a calibration card facing themself 0,5-1 m away from their face, in the gaze direction that the user wants to collect the most accurate data.

2.3 Description of the different metrices

In existing literature, previous studies have used several different metrics to measure gaze concentration. This section presents the identified metrices found in the literature review for this thesis.

Gaze concentration can be described as a decrease in gaze movement that occurs when a driver concentrates their gaze on a smaller area in the visual field. This results in the driver spending more time directing their gaze to the forward path (Wang et al., 2014) while decreasing their time on other AOI such as mirrors and more peripheral areas or the road environment.

2.3.1 Percent Road Center

One way to measure gaze concentration for a driver is to calculate the percentage of data within a defined road center area during a set time interval, so called percent road center (PRC). The data can be defined as the percentage of fixations (fixation-based PRC) but also as the raw collected gaze data (gaze-based PRC) (Ahlström et al., 2009).

Gaze based PRC is a simple and robust choice of measure since the data does not need to be sectioned into fixations before analysis (Ahlström et al., 2009). Gaze-based PRC is highly correlated with fixation-based PRC where Ahlström et al. (2009) showed a correlation coefficient of 0,95 between these measures.

The road center area can be defined in multiple ways. Victor et al. (2005) defines a road center point as the most frequent gaze angle and uses a diameter of 16° or a rectangle of 20° horizontal and 15° vertical for the road center area. Ahlström et al. (2009) uses a central gaze-based PRC where the road center point is defined in the same way as Victor et. al. (2005) but using an 8° diameter for the road center area. A

test study performed on a highway, compared different diameters for central gaze-based PRC. The result showed that a diameter of 8° is not optimal for computing PRC and it should be opted for larger angles such as 12°-16° for increased sensitivity (Wang et al., 2014).

PRC is calculated over a moving time window, where the window size can vary. For instance, Ahlström et al. (2009) calculated PRC over a moving window of four seconds, with increments of one second. Victor (2005) used a moving window of 10, 30 and 60 seconds to detect both visual and cognitive distraction using PRC. Using longer time windows for PRC analysis creates issues that should be noted (Victor & Larsson, 2004). When using a 60 seconds time window, even short non-driving related tasks, such as changing settings on the center console will affect the PRC metric one minute after the action has stopped and the driver has resumed driving without visual non-driving related tasks.

For example, Victor and Larsson (2004) showed that the PRC value fell below 58% when using a time window of 60 seconds while the drivers were distracted by a demanding visual-manual task, such as when dialing a phone number or changing console settings. Victor and Larsson (2004) also showed that drivers who are cognitively distracted show a PRC value of above 92%, a threshold limit used to define gaze concentration. Tivesten et. Al. (2019) also found that drivers that developed overreliance during assisted driving during a complete drive, either had a lot of gaze concentration (PRC > 92% over a 60 second moving window), a lot of visual distraction (PRC < 30% over a 4 second moving window), and/or delayed response times to attention reminders.

2.3.2 Percent Area of Interest

Gaze concentration can also be measured as percent area of interest (PAI). In a previous study that analyzed drivers gaze behavior in roundabouts, PAI was used to measure how much the driver was looking at certain areas (Abbasi et al., 2022). In the study by Abbasi et al. (2022) the driver's field of vision was divided into areas such as the windshield, side windows, mirrors, and interior areas such as speedometer, center screen etc. Gazes that fell in the windshield were categorized as on-road glances while the other areas were categorized as off-road.

When measuring gaze concentration, a variant of PAI can be used, called percent on/off-road. It is a simpler measure to apply as there is only one AOI, an estimated forward road. Some eye tracking systems have defined static gaze areas, see Figure 1, that can more easily be used to estimate the driver's gaze direction. This allows the gaze to be classified as for example in the forward on-road AOI, below the forward AOI (e.g., at a mobile phone) or in other off-road AOI (Abbasi et al., 2022).

In cases where static areas are used, the classification of on-road gaze is typically determined by considering all gaze points within the forward windscreen area. It is important to note that this definition of on-road also includes gaze points towards road signs and the roadside area. Furthermore, some eye tracking systems use post-processing tools that enable the definition of a smaller and more precise AOI for on-road (Tobii, 2022). This allows for a more accurate assessment of gaze concentration specifically related to the road ahead.

PAI is calculated much like PRC and measures the percentage of gaze points in the AOI during a defined time interval. The distinct difference between the two metrics is the size of the AIO as the PRC looks at a smaller road center area.

2.3.3 Gaze entropy

Gaze entropy measures are divided into stationary gaze entropy (SGE) and transitional gaze entropy (TGE) and has been used in several studies to measure gaze concentration (Pillai et al., 2022, Shiferaw et al., 2019, Krejtz et al., 2014).

SGE is a measure of the disorder of the gaze distribution in a defined visual area, for a set timeframe. The visual area could be the driver's whole visual field or a smaller defined area. The visual area can either be split into a grid which contains a defined number of equally sized bins, or that the areas are AOI-based, meaning a non-uniform division of the areas size and position (Shiferaw et al., 2019). The level of SGE is a measure of the distribution of the gaze across the bins. The highest SGE is achieved when the gaze is equally divided over the bins and will decrease when the gaze is limited to fewer bins. An indication of gaze concentration can therefore be found by a decrease in SGE (Pillai et al., 2022). SGE is calculated using the Shannon's entropy equation (Shannon, 1948), see Equation 1.

$$SGE(x) = -\sum_{i=1}^{n} p(i) \log_2 p(i) \qquad 1$$

$$p(i) = \frac{\text{Number of fixations in } i^{th} \text{ bin}}{\text{Total number of fixations}} \qquad 2$$

SGE(x) is the entropy value of the set timeframe *x*. *p(i)* is the probability distribution of a fixation in the i^{th} bin see Equation 2. *n* is the total number of bins. To normalize the entropy value, it is divided by the maximum entropy. The maximum entropy is equal to $\log_2(n)$ (Krejtz et al., 2014).

In the studies presented previously in this section, it is shown that the number of bins and thereby the size of each bin affect the result of entropy. Shiferaw et. al. (2018) used a varying number of bins depending on the visual scenario being analyzed, where more and smaller bins were used when the gaze was more centrally distributed (Shiferaw et al., 2018). Similarly, Pillai et al. (2022) evaluated SGE using four different bin sizes and calculated a signal-to-noise ratio (SNR) to compare the accuracy of the metric. The results showed that using the largest number of bins (100 x 100) resulted in the highest SNR, indicating a more accurate result.

The length of the set timeframe varies between previous studies. Shiferaw et al. (2018) used a one-minute timeframe for analysis of centrally distributed gaze scenarios while five-minute timeframes were used for a total visual field analysis, correlated with the number of bins chosen. Another study used a time window of 10 minutes when studying secondary task (Schieber & Gilland, 2008). Furthermore, a

study detecting alcohol-induced driver impairment, used a time window of 20 minutes (Shiferaw et al., 2019).

TGE is a measure of the disorder of the gaze transitions, meaning how randomly the gaze will transit to a new location, given the current position (Shiferaw et al., 2019). The number and size of bins are defined in the same way as for SGE, where TGE then measures the probability of transitions from one specific bin to another bin. A higher TGE means a more random pattern of gaze transitions, which means a lower TGE refers to increased gaze concentration (Krejtz et al., 2014). To calculate TGE, the conditional entropy equation, Equation 3, is used (Shiferaw et al., 2018).

$$TGE(x) = -\sum_{i=1}^{n} p(i)\left(\sum_{j=1}^{n} p(i|j)\log_2 p\,(i|j)\right), i \neq j \qquad 3$$

TGE(x) is the transitional gaze entropy value for a set timeframe *x*. *n* is the number of bins and *p(i|j)* is the transition matrix containing the probability of a gaze point transition from i^{th} bin to j^{th} bin.

TGE calculates complexity in gaze shift relative to the overall spread of gaze, while SGE is limited to calculate the overall spatial dispersion of gaze. This limitation of SGE entails that a change in SGE may depend on the type of task rather than a change in gaze distribution due to the difficulty of the task (Shiferaw et al., 2019). Shiferaw et al. (2019) discusses therefore that since the two metrics are measuring different aspects, a calculation of both SGE and TGE may generate a more complete evaluation of gaze behavior.

One limitation of calculating TGE is the chosen number of bins. If a small data set with few transitions is used with many bins, the values in the rows in the transition matrix may equal zero. Additionally, if *p(i|j)* is equal to one, which occurs when all the transitions in a row go from one specific bin to another specific bin, the inner sum in Equation 3 will be equal to zero (Krejtz et al., 2015). The entropy value will then show a result of no transitions while in fact multiple transitions occurred.

In one study by Shiferaw et al., (2018) both SGE and TGE were measured on sleep deprived drivers. The result showed an increase in SGE for the sleep deprived drivers compared to the rested drivers. In the same study, TGE did not show a notable increase for the sleep deprived condition (Shiferaw et al., 2018).

Another study by Shiferaw, et al. (2019) used gaze entropy to detect alcohol-induced driver impairment. Alcohol affected the gaze behavior by increasing SGE and decreasing TGE compared to when the drivers were not intoxicated. It was concluded that alcohol reduces the top-down regulation of gaze control which leads to a less structured visual scanning behavior, thereby more gaze concentration (Shiferaw et al., 2019).

2.3.4 Standard deviation of gaze

Standard deviation of gaze (SDG) is a measure of the spatial variability of gaze. SDG is a collective name for three different variants for measuring the standard deviation of gaze: standard deviation of radial gaze (SDRG), standard deviation of horizontal gaze (SDHG) and standard deviation of vertical gaze (SDVG). Several studies have used the three variants of SDG to evaluate gaze concentration (Reimer et al., 2012, Li et al., 2018, Wang et al., 2014). However, SDG has been calculated in different ways across studies, either using gaze vectors in a 2D plane, or gaze angles which are divided in horizontal (yaw) and vertical (pitch) angles. A lower SDG means an increased concentration on the road center and thereby an increased gaze concentration.

SDRG is a combination of the horizontal and vertical gaze angle or gaze coordinates, using the Pythagoras theorem. The radial gaze vector is calculated by the square root of the sum of squared horizontal and squared vertical angles (Victor et al., 2005).

A simplified metric of SDRG is SDHG, which only considers the horizontal position (yaw) of the gaze. Conversely, SDVG is exclusively considering the vertical position (pitch) of the gaze.

According to Li et al. (2018) the SDHG metric show a notable change under added cognitive demand, which Wang et al. (2014) likewise has showed. Furthermore, Reimer et al. (2012) discussed that the combined angles of vertical and horizontal gaze dispersion, SDRG, does not result in a more sensitive measure. In the study from Wang et al. (2014) it was shown that SDHG was more sensitive to measuring driver's cognitive workload than SDRG. Additionally, Wang et al. (2014), discussed that SDG is an efficient metric, because it is straightforward to measure, not relying on a defined arbitrary road center area.

Reimer et al. (2012) found that SDHG increased in response to added low to moderate cognitive demand. Nevertheless, this effect was not observed at the highest level of demand. These findings suggest that there was a bottom out effect at the highest level (Reimer et al., 2012). Furthermore, SDHG has in previous studies, only been used to evaluate cognitive demand. Therefore, it is uncertain whether the metric can accurately detect other forms of impairment.

In the presented studies using SDG to measure gaze concentration, different time intervals have been used. The study from Reimer et al. (2012) calculated SDHG and SDVG for three different events (baseline, cognitive demand, and recovery) and used a time interval of 2 minutes for each event. Another study calculated SDHG with a moving window of 20 seconds with an increment of one second for each event. (Li et al., 2018).

2.3.5 Related measures

In addition to the metrics presented in sections 2.3.1-2.3.4 , there are other visual behaviors that can be captured by an eye tracker and used to detect different driver states. These measures are briefly explained in this section with the purpose of presenting measures that, in addition to gaze concentration, can capture different driver states. However, they were not further analyzed in this thesis.

Glance-based metrics, including glance duration, frequency, and total glance time, are commonly used measures when analyzing visual and attentional demands from information systems in vehicles (Victor, 2005). These measures can be used in analyses when comparing the gaze behavior between different AOI in the vehicle. An indication of gaze concentration to the forward roadway can be fewer and longer on-path glances, as shown in (Victor et al., 2005) where the glance duration increases with increased cognitive load.

The glance durations for looking away from the road center and looking at peripheral objects typically varies between 0,6 to 1,6 seconds, while the glance durations when looking at driver information sources such as the speedometer, clock, etc., typically varies between 1,2 to 1,85 seconds (Victor, 2005)

Another measure used to detect different driver states is the driver's blink frequency. A study by Liang et al. (2007) stated that when drivers are faced with a visual task the blink frequency tends to reduce as the eyes need to process more visual information. The opposite was seen when the drivers were faced with cognitive tasks, where the blink frequency increased from 0,31 Hz, which was baseline, up to 0,49 Hz. This indicates that blink frequency can also be used to detect cognitive load. Increased blink frequency has been identified in other driver states as well. Driver drowsiness has also shown an increased frequency and blink duration (Shiferaw et al., 2018). Previous studies show that blink frequency differs between different types of non-related driving tasks but results also vary between individual drivers (Radlmayr et al., 2019).

3 Methods

This chapter presents the methods of data collection and analysis used in this thesis work. Data was collected through a small study conducted on public roads in northern Gothenburg. For the study, a test vehicle and two eye tracking systems were used to collect gaze data, while the participants were subjected to several secondary tasks to increase the cognitive load. The methods of how the metrics were calculated for each system is presented at the end of the chapter.

3.1 Data collection

3.1.1 Participants

Participants were recruited by a voluntary registration sent out to employees at the two departments: Safety Centre and Personal Driving Experience Centre at VCC. The requirements to participate were that they had a T2 license (an internal level of driving education at VCC that allows driving prototype development vehicles) and that they did not need prescription glasses while driving. Six participants were selected. The small number of participants would ensure meeting the time plan of the project while still collecting data from multiple participants. One participant was excluded from the analysis due to issues with the measurement equipment resulting in missing data. All five resulting participants were male, and the average age was 49 years (min = 39, max = 58). On average, the participants held a driver's license for 31 years (min = 23, max = 40) and drove 14750 km (min = 5000, max = 30000) per year.

3.1.2 Cognitive load tasks

The participants were subjected to two secondary tasks to increase cognitive load during the test drive. Both tasks were provided to the participant through a pre-recorded auditory message that explained the task and was thereafter followed by the actual task.

The first task was a 1-back task where a sequence of numbers was read out and the participant was instructed to repeat the previously stated number, see Figure 3. The task consisted of ten numbers between 0-9 presented in a sequence with 2,25 seconds apart (Wang et al., 2014). After a sequence, there was five seconds of pause before the next sequence started. In total there were four sequences, with a duration of two minutes in total.

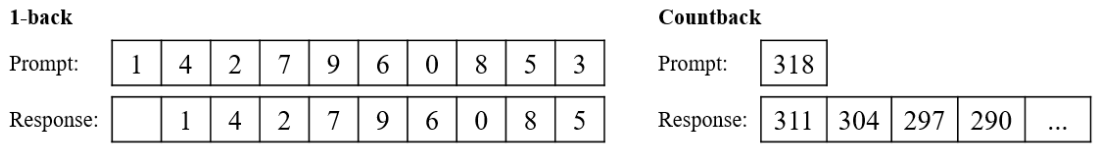


Figure 3: Illustration of the two cognitive tasks. An example of a 1-back task shown to the left, where the participant is instructed to repeat the number that was presented one-back. One sequence is presented but a full task consists of four sequences. An example of a countback task shown to the right, where the participant is instructed to count back towards zero in steps of seven.

The second task was a countback task where a three-digit number was read, and the participant was asked to count backwards to zero in steps of seven (Kountouriotis et al., 2016). After approximately one minute the participant was instructed to stop the task.

The order of the tasks varied between participants. Task 1 and Task 3 were countback tasks while Task 2 and Task 4 were 1-back tasks for three participants, and the opposite for two participants. See Table 1 for the variation of tasks.

Table 1: Variation of order for the different tasks for each participant.

Participant	Task 1	Task 2	Task 3	Task 4
1	Countback	1-back	Countback	1-back
2	1-back	Countback	1-back 1	Countback
3	Countback	1-back	Countback	1-back
4	1-back	Countback	1-back	Countback
5	Countback	1-back	Countback	1-back

3.1.3 Experimental setup

The test vehicle was a modified Volvo XC90 provided by VCC. The vehicle was equipped with a 2-camera system used for research purposes. In addition, Tobii glasses were used for four participants but not for the fifth participant due to technical issues. As the two eye tracking systems operate at different wavelengths, it is possible to use them simultaneously. A portable speaker was mounted to the right of the steering wheel and played the pre-recorded cognitive tasks to the participant while driving. See Figure 4 for the complete experimental setup inside the vehicle.

Additional cameras, two GoPro Hero 11, were placed in the vehicle and were used as reference cameras. The first GoPro was mounted on the right side of the rear-view mirror to capture the forward-facing road and the surrounding environment. The second GoPro was mounted on the dashboard, behind the steering wheel, facing the participant, capturing the participant's face and upper body. The GoPro cameras also collected audio that was used to evaluate the participant's responses to the cognitive tasks.

Two reference markers were placed inside the vehicle used to facilitate post-processing of gaze data for the Tobii glasses (see Section 3.2.1). One marker was placed at the base of the left A-pillar and a second marker was placed to the right of the steering wheel by the center console. The positions were chosen so that at least one marker was expected to always be in view of the Tobii glasses forward facing camera.

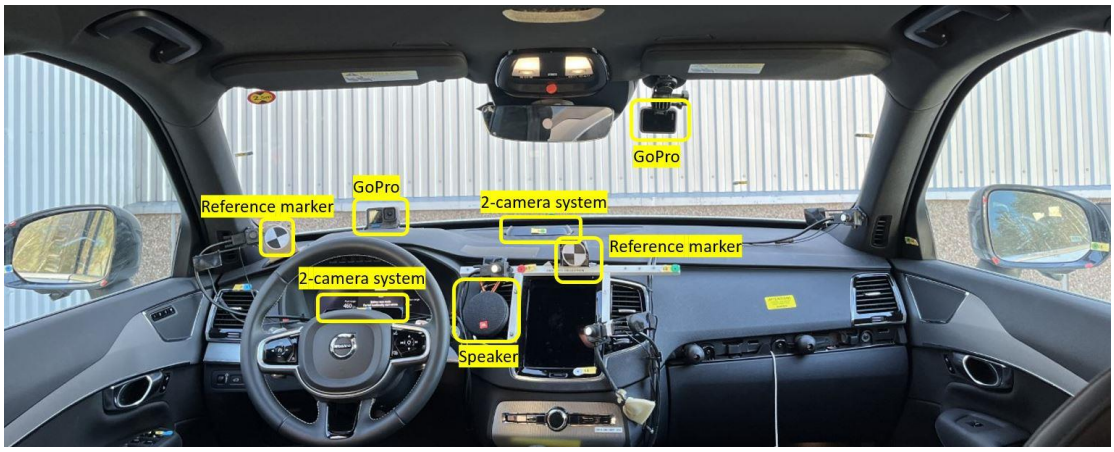


Figure 4: Image showing interior of the test vehicle with complete experimental setup.

3.1.4 Route

The test study was performed on public roads north of Gothenburg, Sweden. The complete test route is shown in Figure 5. The route started with rural road with alternating 1-2 lanes in the direction of travel, with a speed limit of 50-70 km/h before entering the highway with 2-3 lanes in each direction with a speed limit of 80-110 km/h. When reaching the most northern part of the drive (the halfway point), the participants exited the highway and re-entered the highway in the opposite direction and the test drive continued. The trip was approximately 68 km, of which 40 km was driven on the highway. Traffic density during all test drives ranged from low to moderate. The first part of the drive on the rural road allowed the test participants to get familiar with the vehicle before entering the highway and starting the cognitive tasks. For safety reasons, the cognitive tasks were carried out on the highway, to avoid interactions and fewer distractions in the surrounding area compared to rural and city roads.

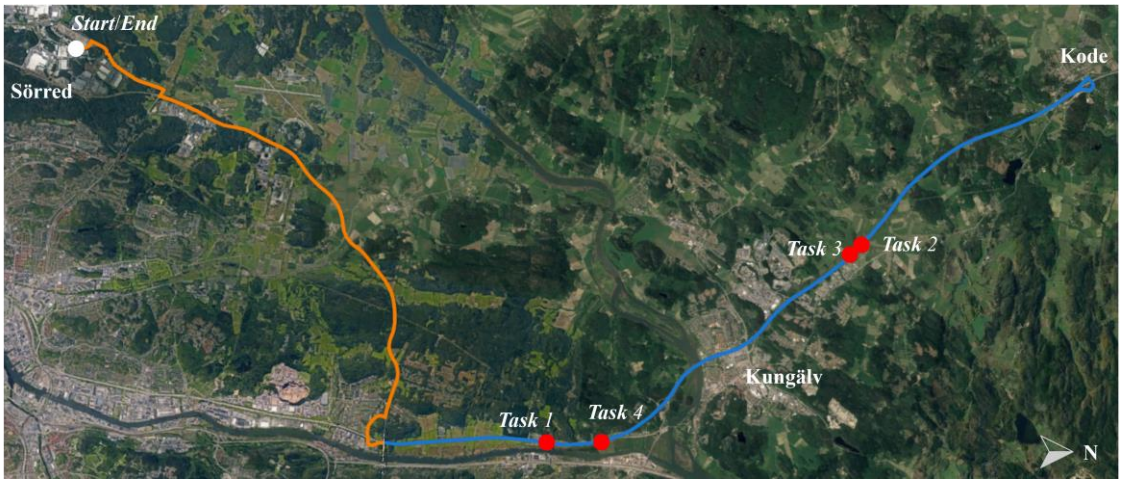


Figure 5: Map showing the test route. The orange line marks country road, the blue line marks highway, the white circle marks the start and end position of the test drive, and the red markings represent the task activation areas.

3.1.5 Procedure

The drive lasted approximately one hour and 45 minutes for each participant. Prior to the study, the participants received written information about the drive. When arriving, the participants were once again briefed about the study aim, data collection, risks, a description of the cognitive tasks, and the driving rules and route. Participants were told to; drive manually (for example not to use any assisted driving systems such as pilot assist), stay in the right driving lane of the highway and only overtake vehicles if the test leader gave permission, and follow the traffic rules. They were also informed that the test leaders would only speak to give directions or, if needed, answer questions related to the test during the drive. The participants could then ask questions about any uncertainties about the test procedure before they signed a consent form. Before the start of the drive, example sounds of the two cognitive tasks were played through the speaker and the participants rehearsed the tasks until they felt confident with them.

If the Tobii glasses were being used, the participants put on the glasses before the drive and were able to change the fitting with purpose of eye alignment with the center point of glasses, as well as comfort. The eye fit was done by changing the nose pads of the glasses, see Figure 2.

Two test leaders were present in the vehicle during the drive. One test leader was seated in the front passenger seat and gave instructions about the route and when the participants were allowed to overtake other vehicles. The other test leader was seated in the left back seat, monitored the data collection, and initiated the cognitive tasks.

A small stop was made outside VCC grounds to calibrate and start the measurement equipment. The participants had the opportunity to change position of the seat, mirrors, steering wheel, and the Tobii glasses a final time before the calibration started. Calibration of the Tobii glasses was done using the calibration card. Each participant was then told to blink four times as a marker for synchronization of all camera recordings during post processing.

For similar driving conditions for all participants during the tasks, the areas of activation of the cognitive tasks were predetermined, see red markings in Figure 5. The activation areas were placed during parts of the highway where there was no entry or exit lane present, nor a change of speed within the first minute of the task. When entering the highway, there was approximately three minutes of driving before the activation of Task 1. Between the activation of each task, approximately five minutes of baseline driving occurred. The traffic situation was assessed as safe by the test leader in the back seat before the cognitive task was initiated when entering the activation area. During the task, the participants kept driving in the right driving lane for safety reasons. If the participants were following a lead vehicle that drove at a speed below the posted speed limit, the participant maintained vehicle-following until the task was completed. The test leader allowed the participants to overtake once the task was completed.

3.2 Data analysis

In this section, the steps in post-processing of the Tobii glasses as well as chosen thresholds for similarity and confidence of the data for the two systems are presented and motivated. Furthermore, methods for calculations of gaze concentration metrics on the data for both systems are explained in detail.

3.2.1 Post-processing of data from the Tobii glasses

Post-processing of the eye tracking data recorded by the Tobii glasses was done in Tobii Pro Lab, a visual user interface made for analysis of recordings from Tobii glasses (Tobii, 2022). The output gaze data uses information from the collected video from the integrated front facing camera. The gaze data is expressed in a local coordinate system, called Head unit coordinate system (HUCS), where the origin is in the center of the front facing camera, see Figure 6. To be able to analyze the data in a coordinate system that does not move, the data needs to be post processed. By selecting a still frame, hereby referred to as snapshot, from the video recording, Tobii Pro Lab uses a built-in Real World Mapping software, so-called assisted mapping, to map the gaze data on to the snapshot.

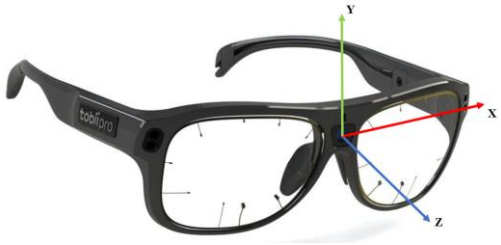


Figure 6: Head unit coordinate system (HUCS) for the Tobii glasses, (Tobii AB, 2023).

The assisted mapping algorithm uses the snapshot from the video recording to place the gaze data points from the video frame on to the snapshot, and thereby expressing the gaze point in a coordinate system based of the snapshot. A similarity score between 0-100% is created for each point, showing the similarity between the snapshot image and the frame image from the video recording. The similarity score therefore represents the amount of information that the assisted mapping could base the mapping upon. A low score indicates that the point's placement is based on less information. Points that had a score of less than 60%, as well as points that were not detected at all by the assisted mapping, were placed using manual mapping. Using a score higher than this resulted in many data points needing manual mapping but that could be accurately placed by the assisted mapping algorithm. Manual mapping was done by observing the gaze point from the video frame, and manually placing it on to the snapshot. The output gaze data from the assisted mapping are expressed in the unit pixels and are expressed in the relation to the pixels in the chosen snapshot the data was mapped upon.

As light and road conditions may differ in the two driving directions during the test drive, two snapshots were used for each participant. A snapshot from the north bound stretch of the highway was used for mapping data points from the first stretch of the

drive, and a snapshot from the south bound stretch was used for mapping data points from the second stretch of the drive.

For extraction of data for analysis within an AOI, a manual AOI polygon tool in Tobii Pro Lab was used for each snapshot to define the road area, see Figure 7. Gaze points that were within the participants' field of view but that don't fit in the snapshot were given designated areas in the snapshot used for manual mapping and can also be seen in the figure. This includes gaze points to the LSM, LE, RSM, and RE, areas that are not found or partially hidden in the snapshot.

Figure 7: Snapshot from Tobii Pro Lab. AOI is showed in purple and is classified as the area for on-road. Pre-defined areas for mirrors and exteriors are shown in red and yellow.

Data points that were not registered by the Tobii glasses, such as when the participant was blinking or when the system for some other reason did not register any gaze direction, were regarded as missing data. Occurrence of these instances were seen during the post-processing as well as in the output data.

3.2.2 2-camera system

The video data recorded by the 2-camera system generated head and eye tracking output signal values, including head position and gaze direction expressed in both angles and AOI. The processed data was measured in the local coordinate system of the vehicle, with the origin at the midpoint of the rear axle.

Each data point has an assigned confidence level ranging from 0-100%, indicating the level of certainty that the correct gaze angle had been determined. A threshold was set where data points with less than 10% confidence were deemed too uncertain and were set as invalid. Data that would indicate that the participant was blinking, or their head was not found by the system (yaw and pitch of 0° and a confidence level of zero) would therefore also be classified as invalid data. The chosen confidence level was selected based on manual analysis consisting of comparison between the video recording and data.

3.2.3 Calculation of metrics

Data synchronization and analysis
Localization of the four blinks that each participant had made in the beginning of the test drive was found in the video recordings from the Tobii glasses, 2-camera system, and GoPro cameras. This allowed synchronization of the video recordings and collected data. A comparison was made between the eye movement of the participants observed in the videos and the data from the 2-camera system to determine if the gaze data showed similarities to the true gaze dispersion.

Events
Ten events were defined along the test drive: Baseline 1-6 and Task 1-4, named in chronological order. The events' start and end points were determined by manual analysis of video and audio data. The 1-back task started when the participant heard the first number of the first sequence, and the countback task started when the three-digit number was presented. The end point for both tasks was after the participant had answered verbally the last number. Baseline 1, 2, 4, and 5 had the same duration as the tasks that occurred after the baseline, while Baseline 3 and 6 had the same duration as the task that has occurred before the baseline. The baseline segments were chosen to mimic the traffic environment of their corresponding task. Takeovers or periods that had higher traffic density were, if possible, avoided in the baseline events.

Percent road center
The most frequent gaze angle was calculated from the start of the first event to the end of the last event for each participant. This created a single most frequent gaze angle representing the entire driving session. For each participant wearing the Tobii glasses, two most frequent gaze angles were calculated, one for each snapshot used in post-processing. The angle was defined as the most frequent gaze angle between the start of the first event to the end of the last event for each of the snapshots.

A diameter of 16° centered around the most frequent gaze angle was used for the PRC calculations in accordance with Wang et al. (2014) to define the road center area for the data recorded by the 2-camera system. As there was to the authors' knowledge no known way of converting the unit from angle to pixels for the Tobii glasses data, a ratio estimation from the gaze points was made where 16° is assumed to be equivalent to 700 pixels. This ratio was calculated using the distance between the most frequent gaze angle and the center of the LSM.

The calculations were made using a moving time window of four seconds with increments of one second to determine the next time window, to compare the PRC over time. This was done for PRC > 92% as well as for an in-depth analysis comparing only one baseline to one task. The PRC was also calculated as one value for each event and participant.

Percent area of interest
For the PAI (on/off-road estimations) all gaze points through the center front windscreen were used as a rough estimate of on-road for the 2-camera system, see Figure 1 that shows the center front area. Gaze points within the defined AOI for each snapshot, see Section 3.2.1, was classified as on-road for the Tobii glasses. The PAI was calculated as one value for the entire event. For an in-depth analysis calculations were made using a moving time window of four seconds with increments of one

second to determine the next time window, to compare the percentage of on-road over time.

Gaze Entropy

SGE was calculated using Equation 1. The chosen AOI was the visual field of the driver. For the 2-camera system data this implied a range of 120° in yaw and 80° in pitch. From analysis of the gaze data, no gaze points were distributed outside of these limits. The size of the snapshots used in post-processing for the Tobii glasses varied slightly in size. Therefore, the largest size of resolution out of all snapshots was used, resulting in 3744 pixels in yaw and 1879 pixels in pitch. The stationary distribution matrix, *N*, was created using a 2D histogram of the gaze data, often referred to as a density plot. Multiple different sizes of the bins used in the histogram were compared to each other before a final bin size was chosen. *N* was used to compute the probability matrix of the stationary distribution, *p(i)*, describing the probability of a fixation in each bin. *p(i)* was calculated as the number of fixations in the bin divided by the total number of fixations. In certain instances, *p(i)* will result in the value zero from the probability matrix. As log2(0) is undefined, the assumption that 0*log2(0) is equal to zero was applied.

SGE was calculated as one value for each event when comparison of all participants was made. For an in-depth analysis, the same calculations were done but over a moving window with a time window size of ten seconds. These calculations did not use any increments of time, but rather used a new sample of the next ten seconds of data points for the next time window.

The same stationary distribution matrix, *p(i)*, calculated for SGE was used for TGE. The transition matrix *p(i|j)* was calculated using the following steps:

1. A sequence of the bins representing the order in which the gaze points occurred chronologically was created.
2. The frequency of transitions between specific bins in the sequence was counted to create a frequency matrix.
    - The bin where the gaze started represents the row number of the matrix, the bin that the gaze transitioned to represents the column number, and the value of the matrix in that specific row and column represents the frequency of transitions.
3. The frequency matrix was normalized to get a probability distribution.
    - The matrix was normalized for each row separately. Each value in a row was divided by the sum of transitions for that specific row.
4. The values from probability distribution were put in the transition matrix.

Using an example, see Table 2, a 3x3 transitional matrix has been created. In bin (1,2) the value is the probability that the gaze transitions from bin number 1 to bin number 2. The diagonal, for example bin (1,1), represents the gaze staying in the same bin. Since this is not defined as a transition, these probabilities are not regarded in the TGE calculations.

Table 2: Example of a transitional matrix

i	*p(i|j)*		
	1	2	3
1	-	0.2	0.8
2	0.7	-	0.3
3	0.5	0.5	-

TGE was calculated as one value for each event when comparison of all participants was made. For an in-depth analysis the same calculations were done as for SGE, using a moving window with a window size of ten seconds. Lastly, the calculated SGE and TGE were normalized, by dividing the calculated entropy with the maximum entropy value. The maximum entropy value is equal to log2(number of bins).

Standard deviation of gaze

Pythagoras theorem was used to calculate the radial gaze vectors used for the calculations of SDRG, see Equation 4.

$$Radial\ gaze\ vector\ = \sqrt{yaw^2 + pitch^2} \qquad 4$$

Calculation of SDHG was done by using the yaw angles of the event, and for SDVG by using the pitch angles of the event. These calculations of SDG resulted in one value for each event for each of the three metrics. To analyze the change of the driver's gaze during each event, the same calculations were done but over a moving window with a window size of 10 seconds. Like SGE, these calculations did not use any increments of time, but rather used a new sample of the next ten seconds of data points for the next time window.

4 Results

4.1 Task performance and data quality

A large amount of data was collected by each eye tracking system, but not all was determined as valid. In this section the type of data and amount that was used in the analysis is presented. All the data except the data from the Tobii glasses data for Participant 1 during Baseline 3-6 and Task 3 and 4, were possible to analyze. This was due to that no data was recorded by the Tobii glasses. These events were therefor removed from further analysis for this specific participant. These events are also disregarded in the scatter plots, showing comparison between the two eye tracking systems.

4.1.1 Task performance

The participants performance of the two cognitive tasks showed a response accuracy of 95% for the 1-back task, and a response accuracy of 98% for the countback task. Consequently, the participants accurately performed the tasks while driving on the selected roads.

During certain events, the participants maintained in vehicle-following due to slow vehicles ahead resulting in slightly different driving conditions for these events. The traffic amount and environment conditions for each event and participant is presented in Appendix AEnvironment conditions during the test study .

4.1.2 Data with low quality

For the 2-camera system data, the percentage of data points used for the analysis is presented in Table 3. The percentage of data points that was classified as invalid due to being below the 10% confidence threshold are presented in the table as well. These data point also included missing data. Results shows that an average of 10% of all data points could not be used.

Table 3: Percentage of gaze points above threshold level (T) and gaze points below the threshold level (X) for each participant (P) and event (B = Baseline, T = Task).

	B1		T1		B2		T2		B3		B4		T3		B5		T4		B6	
P	**T**	**X**	**T**	**X**	**T**	**X**	**T**	**X**	**T**	**X**	**T**	**X**	**T**	**X**	**T**	**X**	**T**	**X**	**T**	**X**
1	94	6	93	7	90	10	89	11	94	6	96	4	90	10	96	4	96.4	3.6	94	6
2	89	11	91	9	89	11	90	10	82	12	90	10	82	12	91	9	89	11	91	9
3	92	8	86	14	91	9	85	15	90	10	93	7	83	17	77	23	80	20	87	13
4	96	4	90	10	93	7	95.4	4.6	96	4	92	8	90	10	92	8	89	11	94	6
5	94	6	83	17	93	7	83	17	89	11	89	11	83	17	90	10	83	17	86	14

The percentage of gaze points from the Tobii glasses data that was determined using assisted mapping by Tobii Pro Lab is presented in Table 4 as well as the percentage of manually mapped gaze points. Two types of gaze points are included in the percentage for manually mapped gaze points; assisted mapped gaze points that were below the similarity threshold, and gaze points that the assisted mapping algorithm could not place. Results shows that an average of 27% of all data points were manually mapped and that 5% of all data points were missing.

Table 4: Percentage of assisted mapped gaze points (A), manually mapped gaze points (M) and missing data (X) for each participant (P) and event (B = Baseline, T = Task).

	B1		T1		B2		T2		B3		B4		T3		B5		T4		B6	
P	A / M	X	A / M	X	A / M	X	A / M	X	A / M	X	A / M	X	A / M	X	A / M	X	A / M	X	A / M	X
1	70 / 25	5	49 / 43	8	30 / 64	6	46 / 48	6	42 / 54	4	- / -	-	- / -	-	- / -	-	- / -	-	- / -	-
2	68 / 28	4	66 / 30	4	72 / 23	5	77 / 19	4	76 / 19	5	74 / 22	4	77 / 19	4	71 / 25	4	64 / 31	5	78 / 18	4
3	85 / 9	6	63 / 28	9	71 / 22	7	77 / 14	9	80 / 15	5	72 / 22	6	77 / 16	7	88 / 7	5	74 / 17	9	79 / 17	4
4	78 / 19	3	81 / 13	6	62 / 35	3	67 / 30	3	72 / 25	3	66 / 30	4	61 / 34	5	65 / 28	7	37 / 53	10	58 / 37	5

Comparison between data and video was done to evaluate the two eye tracking systems' interaction with each other. The data from the Tobii glasses showed no sign of deteriorated signal quality when the 2-camera system was used in parallel. In the video data from the 2-camera system, reflections from the IR LEDs could be seen in certain instances in the Tobii glasses. This led to a decrease in the level of confidence of the gaze angles when a reflection covered the participants pupil, see Figure 8.

Figure 8: Video images from the 2-camera system showing the reflection in the Tobii Glasses from the IR LEDs.

4.2 Calculated metrics

The results from the metrics described in Section 3.2.3 are presented in this section. The results include these metrics for all participants during all applicable events.

4.2.1 Percent road center

Firstly, the PRC was calculated for a moving window of four seconds. The result presented in Figure 9 shows the percentage of PRC for each four seconds period during each event that is above 92%, the threshold for gaze concentration. A value of 0% means that no four seconds period had a PRC above 92%, meaning that no gaze concentration occurred during the event.

Comparing the two systems, it is the same order of magnitude of the PRC calculated from the two systems. In addition, for Participant 2, 3, and 4, the trend through the events is similar in the two plots in Figure 9. PRC is lower during the baseline events compared to the task events, indicating that the driver is concentrating the gaze towards the road center more during the tasks. However, the change in magnitude between the different events is not equal comparing the two systems. PRC calculated from the Tobii glasses shows slightly larger changes between the events. In contrast to

Participant 2, 3, and 4, Participant 1 did not show the same trend for the two systems. The PRC can be seen decreasing from Baseline 1 to Task 1 for the 2-camera system but has the opposite results for the Tobii glasses. The expected outcome of an increased PRC during the task compared to the previous baseline was observed in 11 out of 20 (55%) instances for the 2-camera system and in 10 out of 14 (71%) instances for the Tobii glasses.

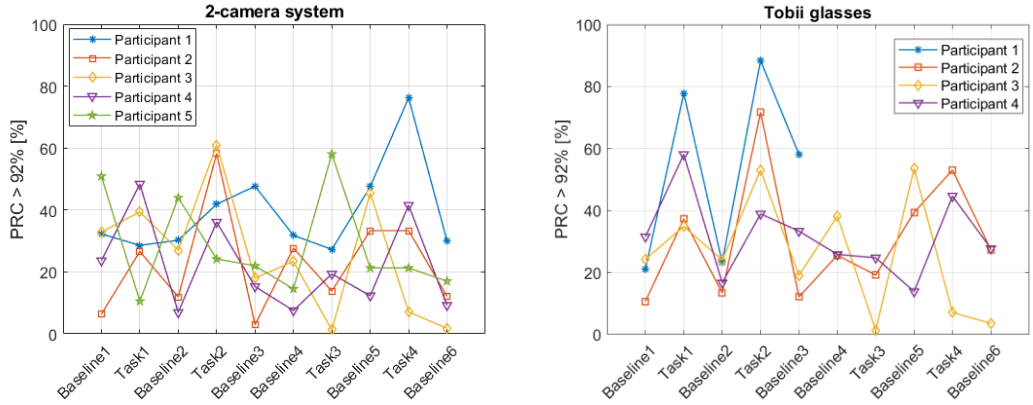


Figure 9: Percentage of PRC above the gaze concentration threshold of 92%.

The events in the two plots in Figure 9 are illustrated in a scatter plot, see Figure 10. If PRC values for each event were equal for the two systems, all data points would be placed diagonally. Most points are placed above the diagonal which shows that PRC > 92% calculated based on data from the Tobii glasses are higher than from the 2-camera system. There is no clear clustering of the baselines or tasks for any participant.

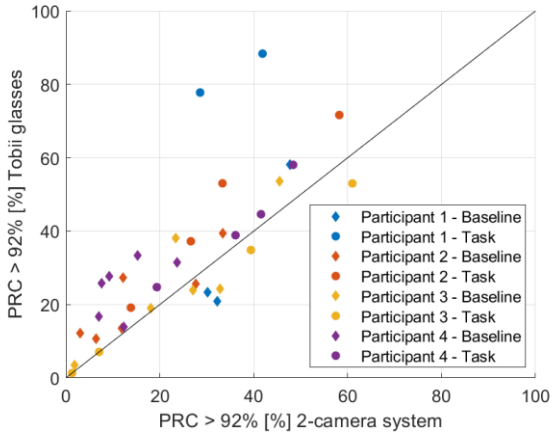


Figure 10: Scatter plot of the percentages of PRC above the 92% threshold for both systems. Data points representing the baseline are marked with a diamond and data points for the task are marked with a circle.

Secondly, the PRC was calculated as a total percentage of all data points within the road center area for each event. As can be seen in Figure 11, the PRC tends to vary between the span of 50-80% for the 2-camera system and 70-90% for the Tobii glasses. The expected outcome of an increased PRC during the task compared to the previous baseline was observed in 10 out of 20 (50%) instances for the 2-camera system and in 10 out of 14 (71%) instances for the Tobii glasses. Comparing the two

systems, they are not following the same trend through the events, to the same extent, as they did for the PRC above 92% presented in Figure 9.

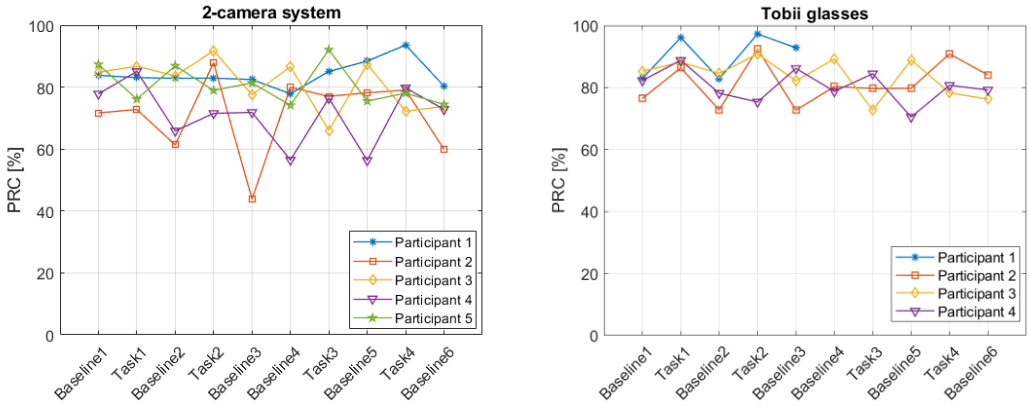


Figure 11: Total PRC for each event.

The scatter plot for the total PRC shows that the values calculated from the Tobii glasses shows higher values compared to the 2-camera system as all points are placed above the diagonal line, see Figure 12.

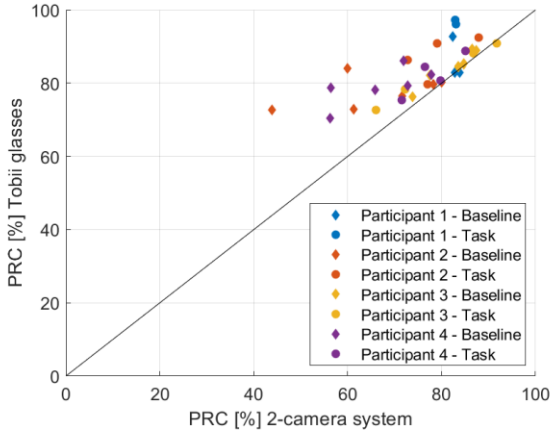


Figure 12: Scatter plot of the total PRC for both systems.

4.2.2 Percent area of interest

The PAI results are presented in Figure 13 and represent the PAI for each event and each participant. PAI for the Tobii glasses shows notable lower values ranging from 25-85% while the results from the 2-camers system range between 75-85%. The magnitude between the different events was also remarkably larger for the Tobii glasses than for the 2-camera system. The expected outcome of an increased PAI during the task compared to the previous baseline was observed in 13 out of 20 (65%) instances for the 2-camera system and in 9 out of 14 (64%) instances for the Tobii glasses.

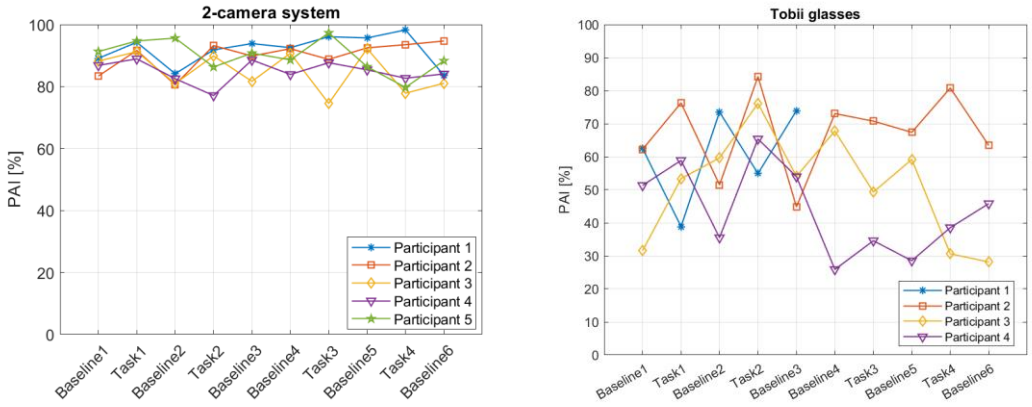


Figure 13: PAI for each event.

The difference in PAI between the systems can also be seen in the scatter plot, see Figure 14. Results show a cluster of PAI for the 2-camera system while a larger spread of PAI for the Tobii glasses.

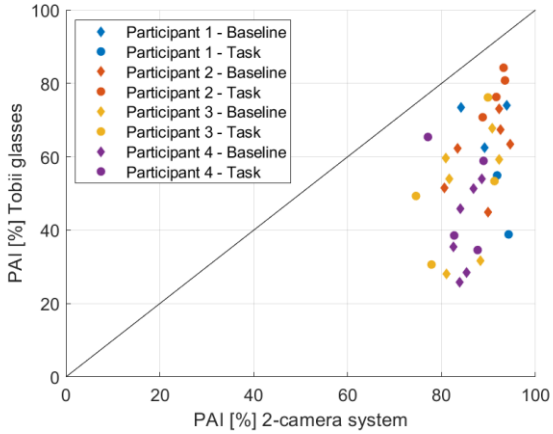


Figure 14: Scatter plot of the PAI for both systems. The 2-camera system is on the x-axis and the Tobii glasses is on the y-axis.

4.2.3 Stationary gaze entropy

To determine the number of bins used in the histogram to calculate SGE, comparison between 11 different bin sizes were performed, see Figure 15. Results for the 2-camera system showed that for bin sizes from 25x25 up to 70x70 the difference in SGE between the baseline and the task remains similar, except for 40x40 for the Tobii glasses. An assumption was made that the choice of bin size was unlikely to have a considerable impact on the comparison of SGE values between the events, within this range. Based on this assumption, a bin size of 30x30 was selected for further analysis, see Figure 36 in Appendix B Density plots with grid size.

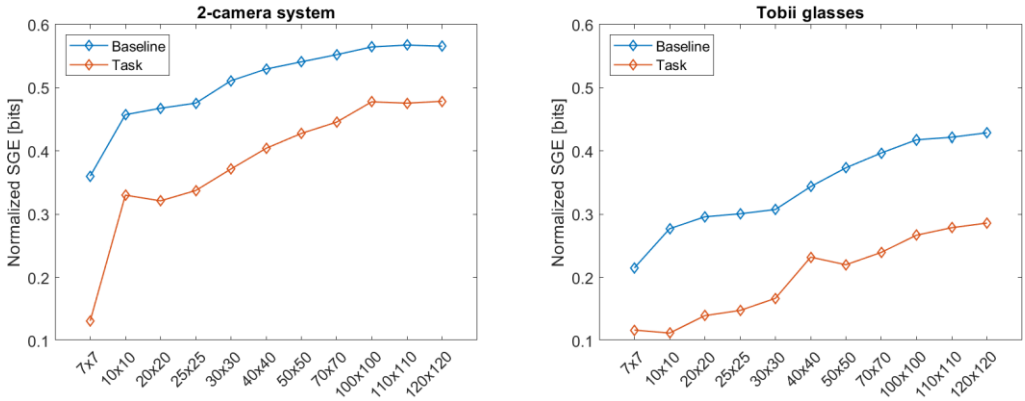


Figure 15: Comparison between different grid sizes with 2-camera system to the left and Tobii glasses to the right.

A pattern of lower SGE during task execution in comparison to the previous baseline was observed in 13 out of 20 (65%) instances for the 2-camera system and 10 out of 14 (71%) for the Tobii glasses, see Figure 16. A low SGE indicates more gaze points within the same bin and can be an indication of gaze concentration. The range of SGE values tends to be the same for both systems, with the 2-camera system showing slightly higher values for some events, for example Baseline 3, 4 and 5. Although they are within the same range, the calculated SGE values can be seen varying in pattern between the two systems for Participants 1, 3, and 4. For example, Participant 4 Baseline 4 show a higher SGE than Baseline 2 in the left plot in Figure 16 compared to the right plot where the opposite can be seen.

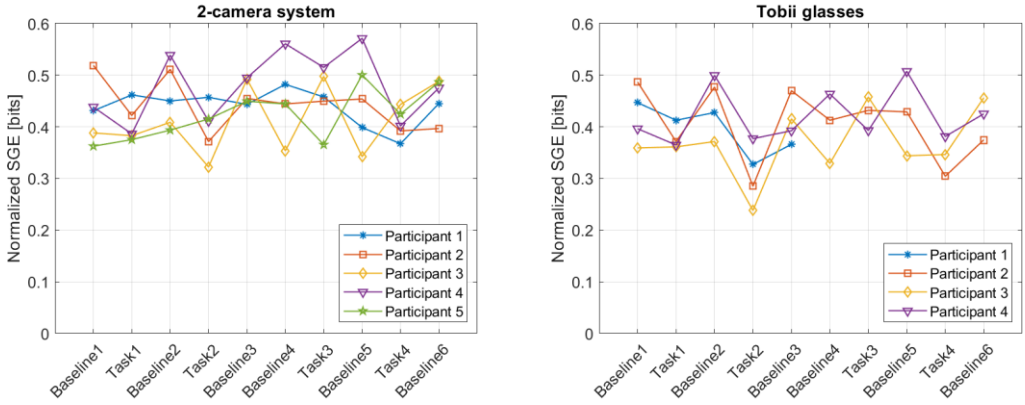


Figure 16: SGE for each event.

A scatter plot of the SGE can be seen in Figure 17 showing similar results from the two systems. As stated in previous paragraph, the results for the 2-camera system were higher while still showing the same trends as the Tobii glasses. There was no distinct difference between results in baseline and task for either system.

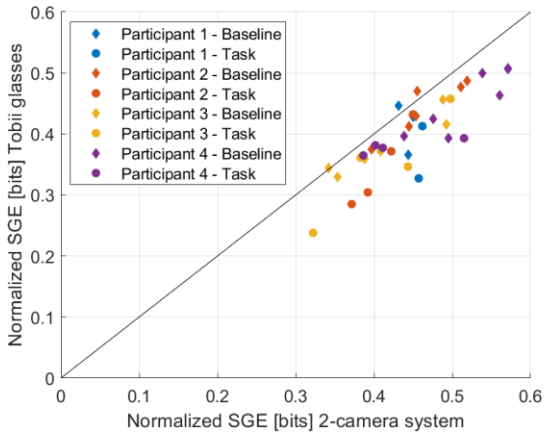


Figure 17: Scatter plot of SGE for both systems.

4.2.4 Transitional gaze entropy

As for SGE, an evaluation was conducted to determine the optimal grid size for calculating TGE. Figure 18 shows a comparison of ten different grid sizes for both systems. For the 2-camera system, the baseline and task events exhibit a similar trend, and the difference in magnitude between the two events remains relatively constant. Conversely, the Tobii glasses show a more arbitrary trend between the two events. The results from both systems indicate that the disparity in TGE between the two events reduces when the grid size exceeds 6x6 bins. A bin size of 4x5 was selected for further calculations, see Figure 37 in Appendix B Density plots with grid size.

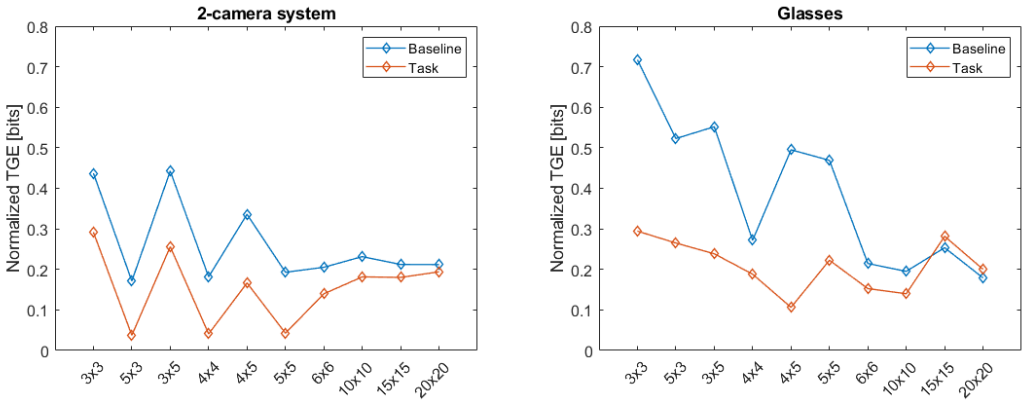


Figure 18: Comparison between different grid sizes with 2-camera system to the left and Tobii glasses to the right.

Much like SGE, a lower TGE represents less transitions of gaze points between the bins and can indicate gaze concentration. The results presented in Figure 19 shows that the magnitude of TGE is similar for the two systems. Participant 5 showed a much lower value in comparison to the other participants for the 2-camera system, and the same could be seen for Participant 1 for the Tobii glasses. Additionally, comparing the two systems, no trends are seen across the different events. As an example, Participant 2 shows for the 2-camera system a lower TGE for Baseline 1 and 2 compared to Task 1, which is the opposite for the Tobii glasses. A pattern of lower TGE during task execution in comparison to the previous baseline can be seen in 9

out of 20 (45%) instances for the 2-camera system and 9 out of 14 (64%) for the Tobii glasses, see Figure 19.

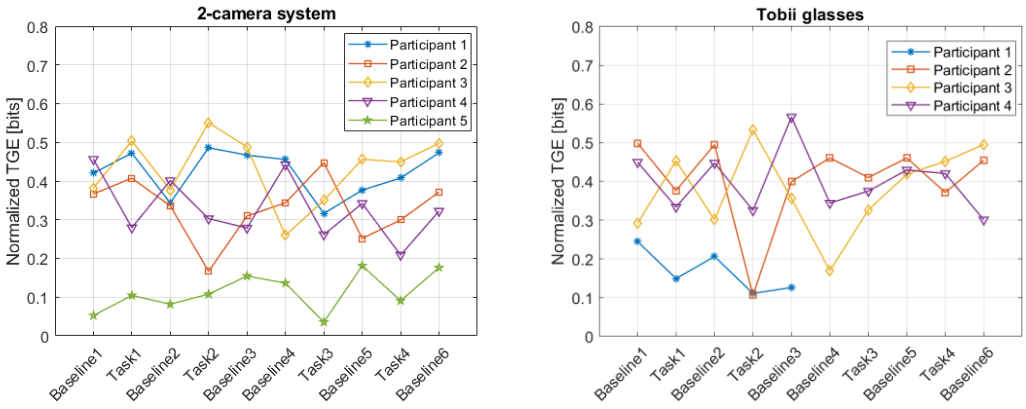


Figure 19: TGE for each event.

The scatter plot for TGE for the two systems is shown in Figure 20, showing a noticeable difference between the four participants. Participant 3 shows similar values of TGE for both systems, while no distinct difference between baseline and task. On the other hand, there was a clear distinction between baseline and task for Participant 4, where most of the baseline values were higher than the task values.

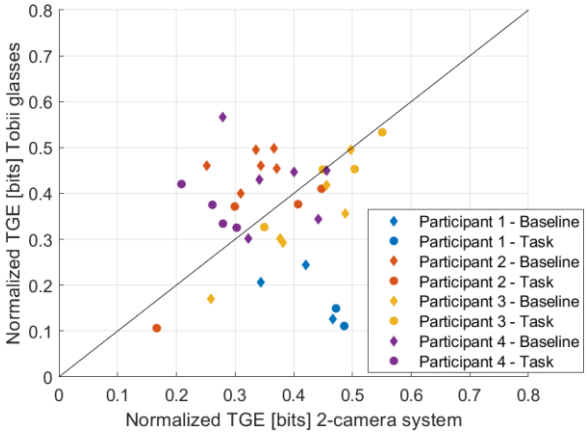


Figure 20: Scatter plot for TGE

4.2.5 Standard deviation of gaze

The results for calculation of standard deviation are presented for SDRG, SDHG and SDVG repeatedly. A pattern of lower SDRG during task execution in comparison to the previous baseline was observed in 10 out of 20 (50%) instances for the 2-camera system and 7 out of 14 (50%) for the Tobii glasses, see Figure 21. As the units measured for the two systems differ, a comparison of magnitudes could not be done. Participant 2, 3, and 4 shows similar trends in results between the two systems, indicating that the systems are similar for SDRG.

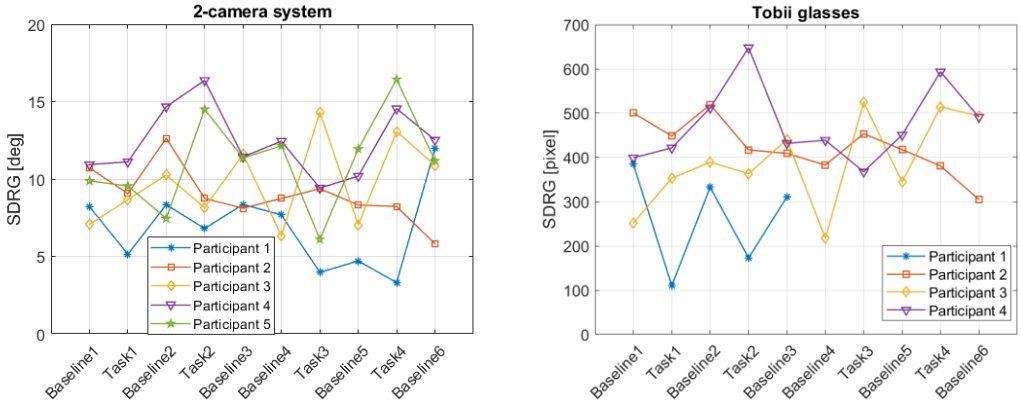


Figure 21: SDRG for each event.

The results from the SDRG scatter plot, see Figure 22, show a similarity in ratio between the systems. Comparison between baselines with tasks shows no clear separation between the two types of events for any of the participants.

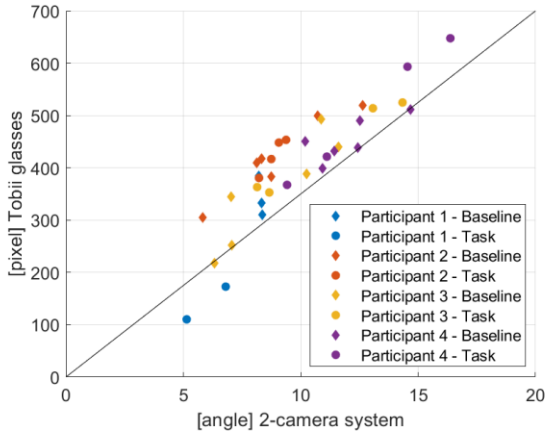


Figure 22: Scatter plot of SDRG

For the SDHG, there were similar patterns for the two systems across the events and participants, see Figure 23. A decrease in SDHG for the task in comparison to the previous baseline was observed for 11 out of 20 (55%) instances for the 2-camera system and for 7 out of 14 (50%) instances for the Tobii glasses.

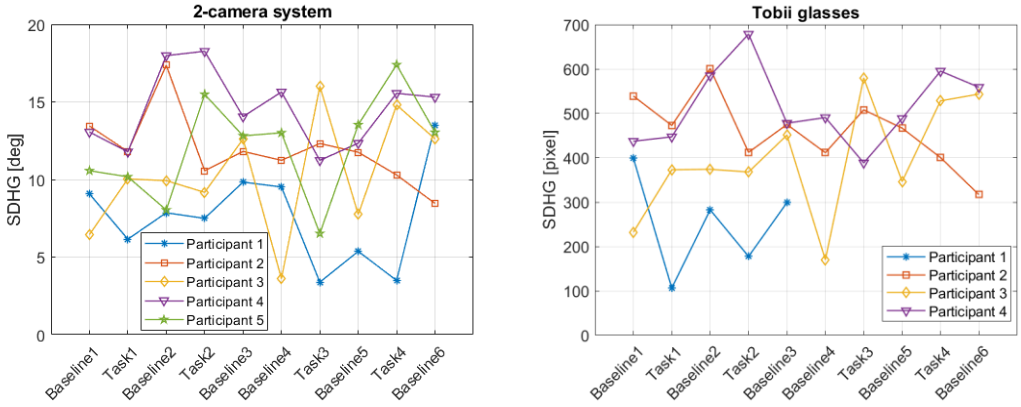


Figure 23: SDHG for each event.

The scatter plot of SDHG for the two systems, see Figure 24, shows a similar result between the two systems as the points are close to the diagonal line. The plot also shows that there was no clear separation between the baselines and the tasks.

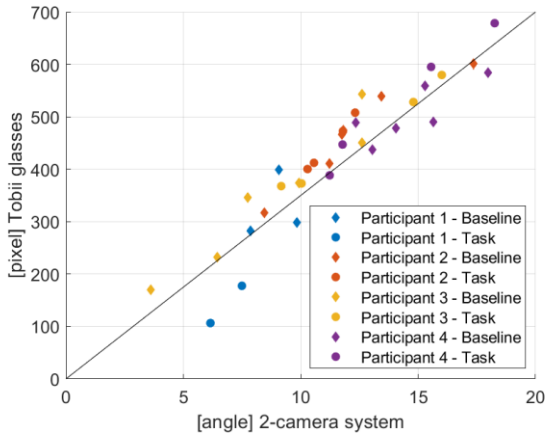


Figure 24: Scatter plot of SDHG

A clear decrease in range can be seen for SDVG in Figure 25 compared to the other standard deviation metrics. The SDVG never reaches above a deviation of 8° and 260 pixels. A decrease in SDHG for the task in comparison to the previous baseline was observed for 14 out of 20 (70%) instances for the 2-camera system and for 10 out of 14 (71%) instances for the Tobii glasses. A majority of the events follow the same trend for both systems with a few exceptions, such as Participant 4 Task 2 for the 2-camera system and Baseline 5 for the Tobii glasses.

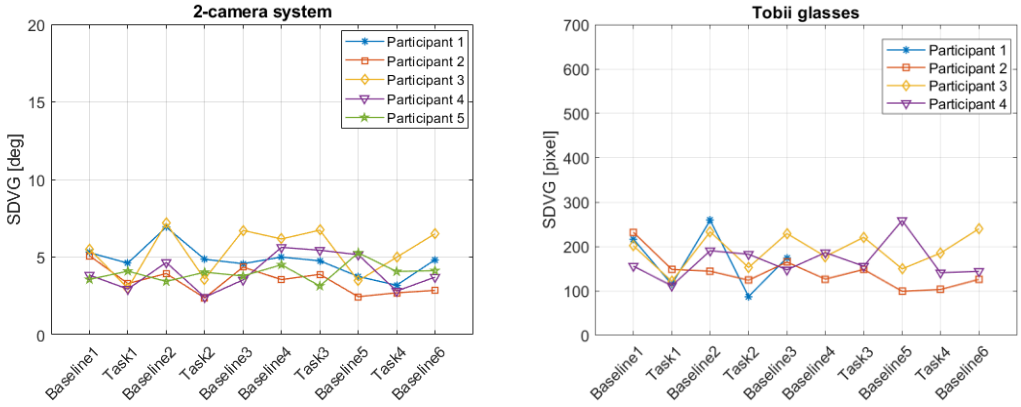


Figure 25: SDVG for each event.

The scatter plot in Figure 26 clearly shows the decrease of range as mentioned in previous paragraph. While the range of values are lower and more clustered for SDVG, a similar pattern can be seen comparing the result to SDRG and SDHG. All three scatter plots follow the horizontal line distinctly.

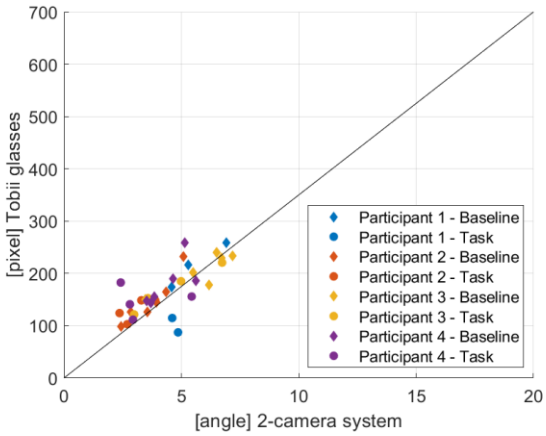


Figure 26: Scatter plot of SDVG

4.3 Comparison between baseline and task regarding gaze concentration

For analysis of how well each metric could capture gaze concentration, Participant 2 was selected for a more in-depth comparison. Using data from both the 2-camera system and the Tobii glasses, comparison was made between one baseline (Baseline 2) and one task (Task 2). During video and data plot analysis, the baseline showed a good example of when the participant had a large distribution of gaze, and the task showed a good example of when the gaze distribution decreases and centers around the road center area. The analysis of this specific segment indicated that the driver exhibited gaze concentration during the task.

Figure 27 and Figure 28 illustrates how the gaze is distributed over the visual field during the two events. These figures shows that the gaze is more concentrated to the road center area during the task, compared to the baseline where there is a larger distribution of the gaze. The color bar demonstrates a higher concentration of gaze point within the road center area for the task event compared to the baseline.

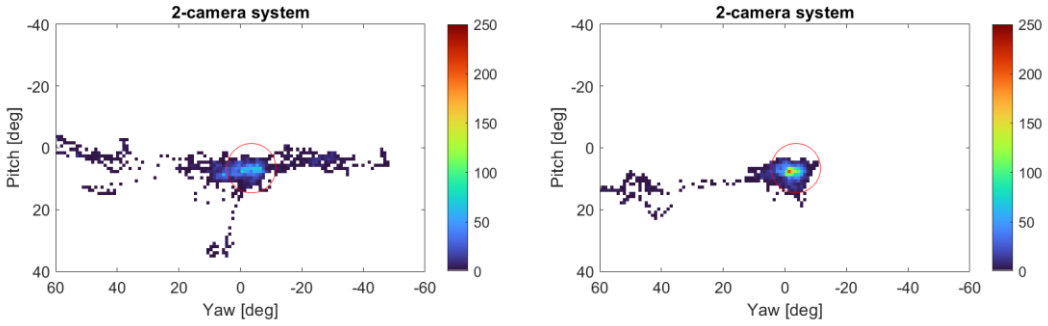


Figure 27: Density plot of gaze angle data from the 2-camera system. The baseline can be seen to the left and the task to the right with the road center area of 16° represented by the red circle.

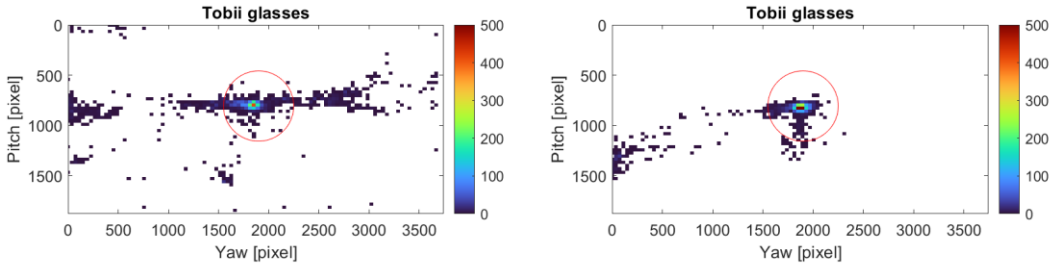


Figure 28: Density plot of the Tobii glasses. The baseline can be seen to the left and the task to the right with the road center area of 700 pixels represented by the red circle.

4.3.1 Percent road center

The PRC represents the percentage of all gaze points that occurred within the 16-degree or 700-pixel road center area, as shown in Figure 27 and Figure 28. The results showed that the percentage of PRC above the 92% threshold for the baseline was 12% and for the task 58% for the 2-camera system. The same analysis for the Tobii glasses resulted in a PRC above the threshold for 13% of the baseline points and 72% for the task points. Comparison of the systems during the baseline show that during the interval of 0-25 seconds, the 2-camera system exhibits a lower PRC (10-60%) in contrast to the Tobii glasses which shows a higher PRC (32-85%). After this they both exhibit the same trend. During the same interval, the 2-camera system shows a gradually increase in PRC, a behavior which is not shown for the Tobii glasses. During the task, both systems showed similar trends with one exception during the period of 10-20 seconds where a decrease in PRC is seen for the 2-camera system but not for the Tobii glasses.

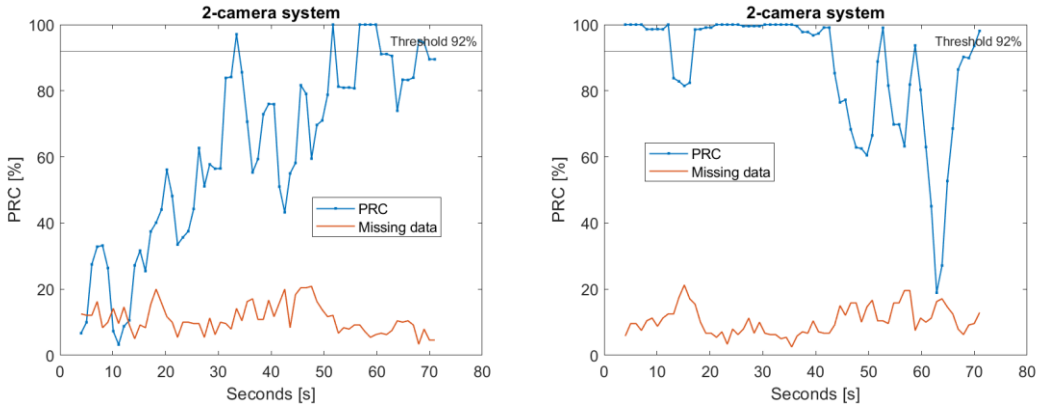


Figure 29: PRC for the 2-camera system. Each point in the figure represents the PRC value of a four second period. Baseline is presented in the left figure and task in the right figure.

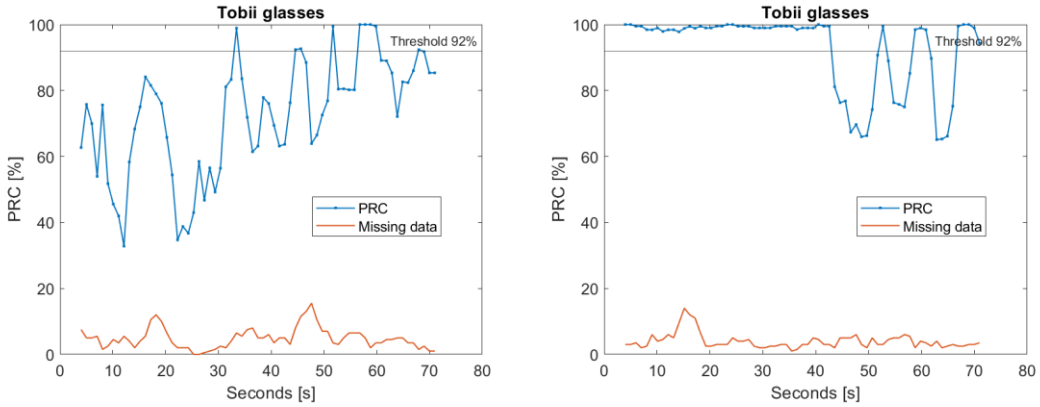


Figure 30: PRC for the Tobii glasses. Each point in the figure represents the PRC value of a four second period. Baseline is presented in the left figure and task in the right figure.

Video analysis of the two events showed that during the baseline, the participant looked more towards oncoming traffic and surroundings in the first half of the event in comparison to the second half. During the task, the participant directed their gaze forward during the first half of the event in comparison the second half where multiple glances were made to the left-side mirror explaining the decrease in PRC.

4.3.2 Percent area of interest

The PAI for the two systems are presented in Figure 31 and Figure 32. A significant difference could be seen between the baseline and the task for both systems. Throughout the baseline, the PAI reached higher values for the 2-camera system, reaching up to 100% for multiple instances compared to the Tobii glasses which never resulted in a PAI above 85%. The same trend can be seen for the task, where PAI never reaches 100% for the Tobii glasses. While the Tobii glasses show a lower PAI, the difference in size between data points are often the same for the two systems.

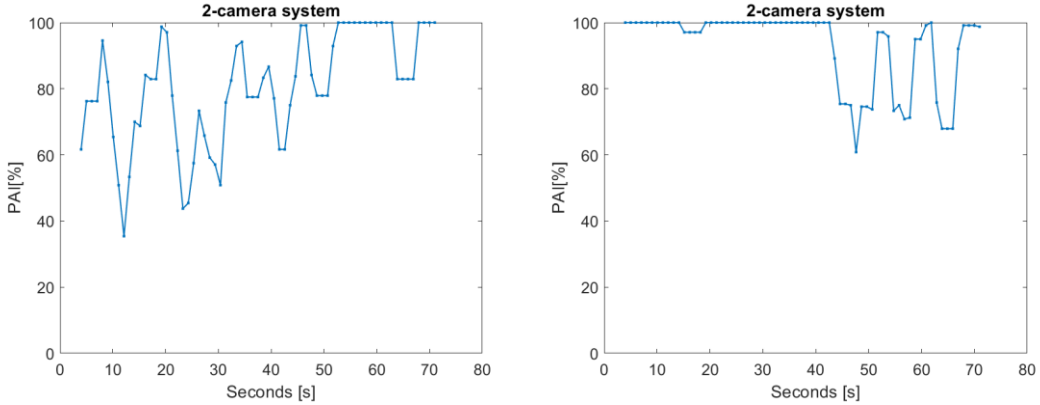


Figure 31: Percentage on-road for the 2-camera system. Each point in the figure represents the PRC value of a four second period. Baseline is presented in the left figure and task in the right figure.

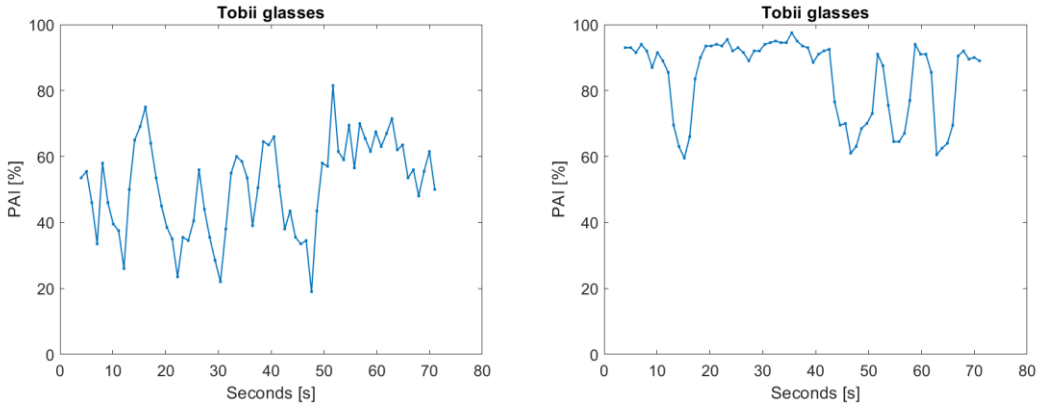


Figure 32: Percentage on-road for the Tobii glasses. Each point in the figure represents the PRC value of a four second period. Baseline is presented in the left figure and task in the right figure.

4.3.3 Stationary gaze entropy

When analyzing the SGE over time, significant differences could be seen between the two events for the 2-camera system, see Figure 33. While the two systems exhibit similar trends during both events, the Tobii glasses have a slightly larger change in absolute values between data points.

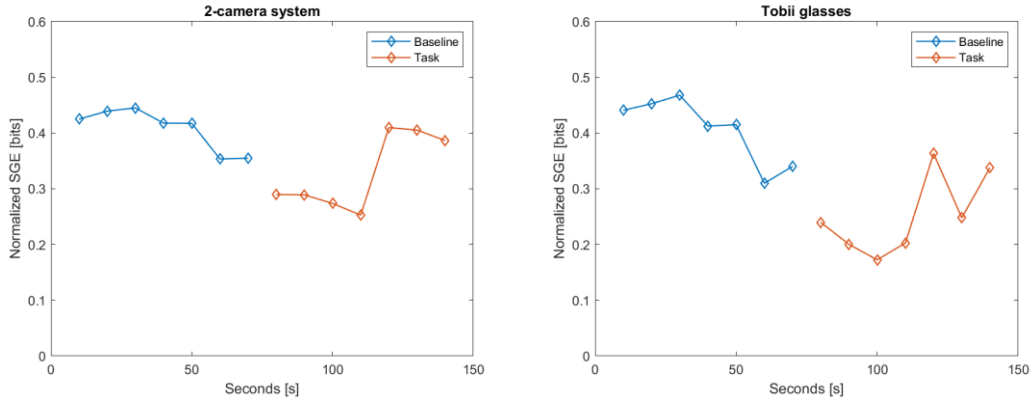


Figure 33: SGE calculated over time, 2-camera system to the left and Tobii glasses to the right.

4.3.4 Transitional gaze entropy

The calculated TGE using a moving time window showed that clear differences could be seen from the two events for the 2-camera system, see Figure 34. A clear difference could also be seen between the two systems, where multiple TGE points reach zero bits for the 2-camera system during both the baseline and the task, while the Tobii glasses fluctuate steadily around a TGE of approximately 0,12 bits. The expected results of a lower TGE during the task was seen for the 2-camera system but not for the Tobii glasses.

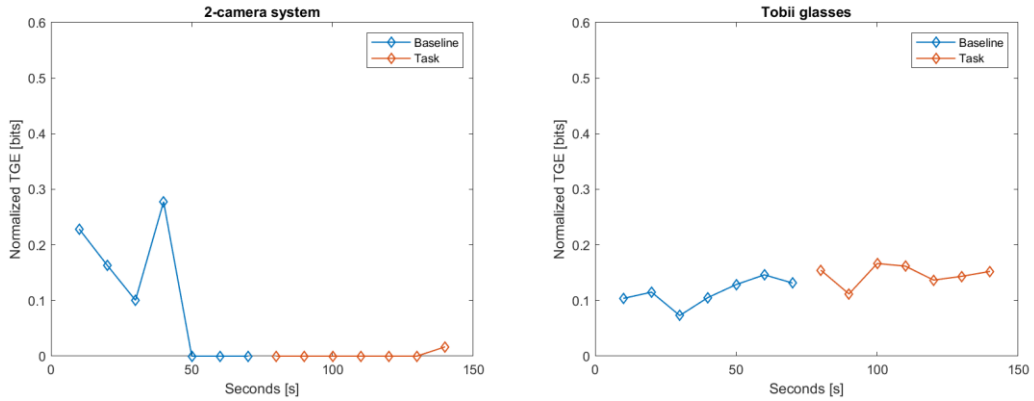


Figure 34: TGE calculated over time, 2-camera system to the left and Tobii glasses to the right.

4.3.5 Standard deviation of gaze

SDRG, SDHG and SDVG were all calculated using a time interval of 10 seconds for the whole event, see Figure 35. For the 2-camera system, similar trends were seen for SDRG and SDHG comparing the baselines as well as the tasks. The same trend was not seen for SDVG. Similar tendencies could be seen for these two metrics for the Tobii glasses, but SDHG show larger differences in results for both the baseline and the task.

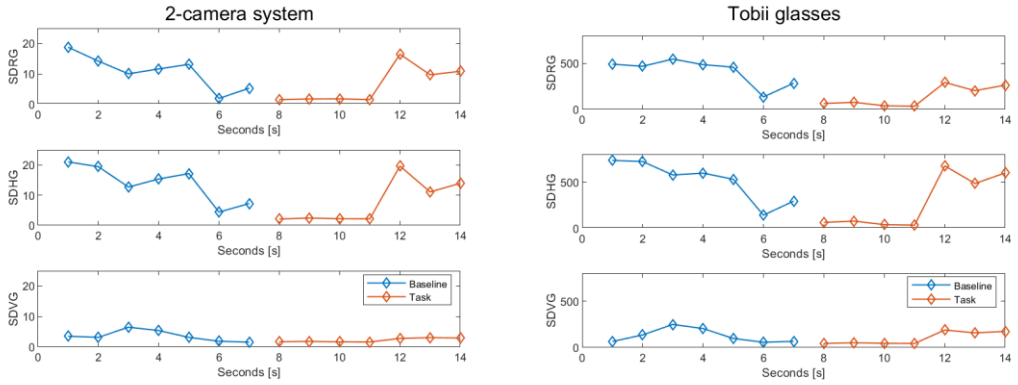


Figure 35: Comparison of standard deviation for the 2-camera system to the left and Tobii glasses to the right.

5 Discussion

Gaze concentration may be an indication of cognitive distraction and other types of driver impairments (for example intoxication and fatigue). In this chapter, gaze data that was collected from the test study using two eye tracking systems; an in-vehicle 2-camera system and Tobii glasses is discussed. The metrics that are commonly used in literature to indicate gaze concentration are discussed and compared in terms of reliability and accuracy. The data availability of the 2-camera system and the Tobii glasses was compared in terms of invalid and missing data as well as how well the different metrics show gaze concentration. Suggestions for future work includes discussion on how these metrics can be applies in real-time eye tracking as well as how larger research studies can explore areas that were limited in this thesis.

5.1 Data availability in different systems and metrics

The 2-camera system's confidence level is the only indication of how accurate the measured gaze data was. As mentioned in Section 4.1.2, a decrease in confidence level, was observed in the data from the 2-camera system when reflections from the IR lights in the Tobii glasses covered the participants' pupils. However, the in-depth comparison of these segments revealed that despite the decrease in confidence, the gaze direction remained accurate for these segments. Hence, the confidence level does not explain why the data was still reliable even though the confidence level was low. This shows that data points with zero confidence could still be used in certain instances, and that excluding all data points below a confidence threshold will lead to inaccuracy of the true gaze direction. As explained in Section 3.2.2, angles that has a yaw and pitch of 0° in combination with a confidence level of zero, was treated as missing data and excluded from the analysis. Failing to exclude these values would then also lead to skewed results.

When averaging the amount of missing and invalid data across all participants and events for the two systems, the 2-camera system classified twice the amount of gaze data as invalid (10%) compared to the missing data from the Tobii glasses (5%). Although the percentage of invalid data was higher for the 2-camera system, the data from the 2-camera system does not need post-processing, a crucial step that was needed before any metrics could be derived from the data from the Tobii glasses. For the Tobii glasses, an average of 27% of the gaze points for each event were manually mapped, indicating that almost a third of the gaze data would have uncertain positioning if no post-processing were done. It can be argued that a similarity threshold of 60% was not high enough. Analysis of the gaze data showed that arbitrary gaze points with a similarity threshold over 60% were still mapped incorrectly, meaning that uncertainty in accuracy of the data remains. Additionally, due to the design of the snapshot, see Figure 7, gaze points that fall within the designated areas for the side mirrors will not show the true positions or transitions of the gaze that occurs. Since the sequenced gaze points are manually mapped to the designated area they will not match in distance to the actual mirror.

5.2 Comparison of trends and magnitudes between the 2-camera system and Tobii glasses

Each metric has been compared in terms of trends and values from the results in Chapter 4. In this comparison, the term "trend" refers to whether the result shows the same patterns, such as an increase or decrease in value for an event. These trends have been compared between the two systems as well as discussed in terms of how well they show indication of gaze concentration.

5.2.1 Percent road center

The calculation of PRC > 92% showed similar trends across the events for both the 2-camera system and the Tobii glasses. However, the scatter plot, Figure 10, showed that PRC > 92% calculated from the Tobii glasses tends to yield higher values compared to the 2-camera system. Moreover, the PRC calculated over the whole event, show values with less of a range when using the Tobii glasses, as seen in Figure 12. The higher values for the Tobii glasses may be due to the different definitions of the road center area used by the two systems.

Furthermore, when comparing PRC between baseline and task in Section **Fel! Hittar inte referenskälla.**, similar results were observed. It was noted that when the driver's gaze was focused on the road center for a longer duration during the task (20-40 seconds in Figure 29 and Figure 30, both systems exhibited a PRC of 100%. This would imply that the road center area captures similar visual field for the two systems. Additionally, variations in PRC could be attributed to the amount of invalid data. The discrepancies observed between the systems during the 10-20 seconds interval of the task, see Figure 29 and Figure 30, could potentially be explained by differences in the amount of missing data during that time.

5.2.2 Percent area of interest

For the results of the PAI calculations, the 2-camera system consistently shows a higher PAI for all participants and events compared to the Tobii glasses, see Figure 13. The reason for this difference is the definition of the on-road area. The 2-camera system uses the entire front center windshield area, and the PAI will therefore not distinguish between gaze points the driver has on the road or on traffic signs, for example. The Tobii glasses on the other hand has a much more narrow definition of the on-road in the snapshot and gives high accuracy when the gaze shifts between defined AOI.

Analysis of the range in Figure 13 shows that the Tobii glasses has a larger difference (25-85%) between events than the 2-camera system has (75-85%). This would indicate that a PAI calculated using a smaller defined area of on-road, would show a greater difference between normal driving and driving with additional cognitive load. Hence, the Tobii glasses is the preferred system for PAI calculations to ensure more reliable indication of when gaze concentration occurs.

Comparing the result of the Tobii glasses in Section **Fel! Hittar inte referenskälla.** and **Fel! Hittar inte referenskälla.**, PAI show similar trends but lower values than PRC. This decrease in value is mostly due to the definition of the road area which

does not capture the horizon above the road, an area close to where the driver frequently places look-ahead-fixations (Lehtonen et al., 2013). However, comparison of the same results for the 2-camera system shows that PAI have higher values than PRC, an expected outcome because of the larger AOI for the PAI.

5.2.3 Stationary gaze entropy

The bin comparison in Figure 15 showed that using different grid sizes on the same data set results in different values of SGE, which is also showed in (Pillai et al., 2022). Another aspect when comparing SGE across data sets, is the fact that the entropy values are normalized with respect to the maximum entropy, calculated by the number of bins used. There is no natural maximum of entropy and hence it can be argued that normalized entropy will not show a representative result when comparing results between data sets without taking the number of bins into consideration (Shiferaw et al., 2019). There is no known threshold of SGE that would indicate gaze concentration, showing the difficulty of establishing how the metric can identify this visual behavior.

Although the same number of bins were applied to both systems, the visual field which the bins were distributed over was different. Figure 36 in Appendix B
Density plots with grid sizeillustrates how the gaze points were distributed for Participant 2, showing a larger distribution of gaze points across the visual field from the Tobii glasses compared to the 2-camera system. The difference in data units from the systems results in gaze points being distributed across the bins differently. This division of data points will affect the SGE results since the percentage of data points in each bin and the number of bins that contain data points differs. The density plots show that a significant portion of the data points was concentrated over a small number of bins. SGE results will change whether these points are counted in a single bin or divided among multiple bins, explaining why the SGE between the systems varies.

Outliers in the data, such as falsely measured gaze points, may contribute to more bins filled with gaze points than the true case. Considering this factor in combination with the difference in bin distribution mentioned in previous paragraph, the slightly higher SGE values for the 2-camera system indicate that the gaze was more evenly distributed over the bins and/or that the gaze points were distributed across more bins. Although higher values, the two systems show similar trends of results for all participants in Figure 16, indicating that there is no distinct difference when applying the metric to the two types of gaze data.

5.2.4 Transitional gaze entropy

The results of the bin size comparison revealed that the chosen number of bins significantly affect the values and range of TGE between the two events. Moreover, the large variety in TGE values results in a challenge comparing TGE across different datasets with different number of bins. The noticeable variations and stochastic trends in the analysis of different grid sizes, particularly observed in the Tobii glasses, suggest that TGE is not a robust metric.

Since TGE is a measure of how often the gaze moves across the defined visual field and to what position the gaze is moving, the placement and size of each bin directly impacts the result. For instance, the road center area being divided into one bin or multiple bins, will change the outcome of the number of transitions between the bins. The conventional definition of TGE in the literature divides the visual field into a grid of bins with equal size (Krejtz et al., 2014). However, an alternative approach for SGE and TGE is to define the bins as different AOI within the visual field (Shiferaw et al., 2019). In this case, the road center area would be represented in one AOI, while for example the side mirror would be defined in another AOI. The results of TGE would then be independent of where the edges of each bin are placed, which can otherwise affect the outcome. Since small gaze transitions are not of interest for this measure, using AOI as bins may yield more relevant results. Furthermore, this definition could serve as a more robust metric when comparing different data sets or data from different measurement systems, as each AOI is customized to the data set.

In Section 4.3.4, the results show that TGE measured for a time window of 10 seconds yielded a value of zero for several measurement points. As explained by Krejtz et al. (2015), having too many bins in relation to the data may lead to situations where no transitions occur to certain bins. This would result in that some rows in the transition matrix sum up to zero. Moreover, when there are few transitions, the gaze may consistently move between specific bins. This situation results in a probability equal to one, as there is no uncertainty regarding the gaze transitions. These two scenarios will result in an entropy value of zero, regardless of the presence of gaze transitions.

5.2.5 Standard deviation of gaze

SDRG and SDHG show the same trend across the events and participants, for both systems, see Figure 21 and Figure 23. The deviation is larger for SDHG compared to SDRG for all events and participants, confirming the same findings as Wang et al. (2014). Furthermore, Wang et al. (2014) found that SDVG is not a sensitive measure for cognitive demand, in line with the results in this thesis. The range of values is smaller and much more clustered than the other measures, indicating that SDVG is not the preferred metric to calculate SDG.

While the Tobii glasses showed slightly higher values for SDRG, the same could not be seen for SDHG where a more equal result was presented between the two systems. Equivalent results could be seen in Figure 33 for all SDG metrics. It can be noted from the in-depth comparison that both SDRG and SDHG show value close to zero when the driver gets gaze concentration, studying the first half of the selected task event. This highlights that both metrics are sensitive measures of gaze concentration.

5.3 Comparison of metrics

While many previous research studies have used the metrics presented in this thesis, they most often have one factor in common. The studies tend to only examine one or two metrics for the same dataset. It is therefore difficult to know if the result of previous studies would have similar findings of gaze concentration if they selected a different metric.

Considering all participants and events, PRC, PRC > 92%, PAI and SGE present similar trends in terms of indicating gaze concentration. Furthermore, for one participant, TGE yielded similar trends as previously stated metrics, but this was not seen for the other participants. This may indicate the one participant was an outlier and that TGE in fact has little similarities with the other metrics.

The SDRG and SDHG showed remarkably similar results, where the results aligned with PRC, PRC>92%, PAI and SGE. However, the similarity in trends were not as clear for the SDRG and SDHG. Additionally, SDVG displayed small variations between the events and does not follow the same trend suggesting that this is a metric that is not sensitive to gaze concentration.

Comparing the metrics regarding the sensitivity of the correct gaze placement, SGE, TGE, SDRG, SDHG and SDVG all consider the gaze points placements and movements in a world coordinate system. Therefore, these metrics are more dependent on the exact placement of all gaze points in comparison to PRC and PAI, which only consider whether the gaze is placed in the defined area or not. Therefor PRC and PAI are also less sensitive to gaze points subjected to variations in the peripheral visual field, which the other metrics mentioned in this paragraph are highly sensitive to.

5.3.1 Application of the metrics in real-time analysis

If the metrics in this thesis are to be used in real-time eye tracking systems to identify gaze concentration, certain conditions must be met. The eye tracking system needs to be capable of collecting, filtering, and computing gaze metrics in real-time while driving. Due to the need for post-processing of data, the Tobii glasses cannot be used as a real-time eye tracking system.

All metrics analyzed in this thesis were calculated based on multiple data points within a set time window. Therefore, the metrics cannot be computed from the instance of start of the drive but rely on data with the same duration as the time window used for analysis. However, the amount of data required for calculating each metric varies.

PRC relies on a defined road center area, which needs accumulated gaze data to calculate the most frequent gaze angle. Due to the definition of the most frequent gaze angles of being calculated for each new participant and each new drive, it would need to be calculated dynamically while driving. Challenges with calculation of PRC include the road design, individual behavior of the driver, and changes of the environment such as weather and road slopes, that all have impact on where the most frequent gaze angle would be placed. In such cases, PRC may be incorrect due to discrepancies between the actual road center and the calculated road center. More extensive research is needed to investigate how sensitive PRC metrics are to different road geometries and how important an accurate classification of the road center is. This would include whether the road center should be fixed or dynamic as well as the shape and size of the road center area.

PAI, however, does not require as much previous collection of gaze data as PRC. If the eye tracking system being used has pre-defined AOI, such as forward center, the

system can define the gaze as on- or off-road. The time span of previous gaze data needed will then be defined as the chosen time window for the calculation of the metric.

When calculating SGE and TGE, longer time windows of collected gaze data is needed to avoid unreliable results due to few data points, as discussed in section 5.2.4.

Standard deviation of gaze metrics (in other words SDRG, SDHG, SDVG) only requires the amount of data equal to the chosen time window of analysis. This makes them fast and reliable metrics in terms of real-time application as the time window can be set and adjusted while also directly applied on the collected gaze points.

The duration of the moving time window depends on the specific metric and the driver state being analyzed. For instance, visual distraction is a transient driver state that requires a relatively short time window to be able to observe a change in the gaze behavior. If using longer time windows, the chance of not capturing the driver state in time may occur. For example, if a cognitive task lasts for 60 seconds, using a time window of the same length would then only capture gaze concentration at the very end of the task. On the other hand, a non-transient driver state such as sleepiness or overreliance in assisted driving gradually increases over time, leading to gradual change in gaze behavior. Driver states that last for longer periods of driving will therefore require the metrics to be applied to longer time windows.

5.4 Future work

As invalid or missing data occurred for both systems, it would be of interest to understand when the loss of data occurs. This thesis did not include an investigation whether both systems experienced loss of data at the same time or occurred arbitrarily. Comparing this type of analysis with collected video data would allow for a better understanding of the eye tracking system’s limitations.

To the authors' knowledge, there is currently no prior research studying the comparative performance of metrics applied to data from different systems. This presents an opportunity for conducting research studies which compare and evaluate various measurement systems using the presented metrics. Furthermore, this study only assessed the metrics' performance during cognitive load, suggesting the need for future research to explore their performance in other types of transient and non-transient states. To draw statistically significant conclusions about the metrics and eye tracking system performance, future research should include more participants resulting in more datasets that can be used for analysis.

An in-depth comparison was only done for one baseline and one task as presented in Section 4.3. Results show both similar and opposite trends compared to the remaining participants, suggesting that the results may vary if other events had been analyzed. To analyze if differences between participants were present in the data, statistical tests should be performed on the data. This should be done to see if any random or individual differences between participants are present and minimize the risk of possibly neglecting fixed and random effects. Application of an in-depth comparison and the use of a mixed-effect model should be done to find outlying participants and results in the metrics.

5.5 Limitations

The test study only included a small number of participants, all employees within two departments at VCC. Therefore, the range of variation among different volunteers (for example age, gender, occupation etc.) were limited, and the selection was not entirely random. The dataset used for evaluating the selected metrics is too small to draw any statistically significant conclusions, and the observed differences could not be generalized to a population of drivers. The test study only subjected the participants to cognitive distraction, and other driver states such as intoxication and fatigue were not tested.

It remains uncertain how accurate the collected gaze data was, or which of the two eye tracking systems that has collected data closest to the true gaze directions. Comparison between metrics and systems should therefore be interpreted with caution. To be able to truly compare the eye tracking systems, a ground truth should be used as a baseline for all metric calculations. This could include collection of gaze date from a high precision eye-tracker used simultaneously.

As stated in Section 3.2.3, the authors are not aware of any method for converting the unit of pixels to angles. No analysis was made to confirm that the assumption that 16° corresponds to 700 pixels was correct. Therefore, a true comparison of PRC using the same road center areas, as well as normalization of the SDG values from the systems could not be made.

A method that could be implemented to compare data more easily from the two systems would be to include the raw data from the Tobii glasses. This data is expressed in gaze direction based on the Head unit coordinate system in combination with gyroscope data, to express the data from the two systems in the same coordinate system. It was found that the gyroscope data was missing for participant, resulting in that this method was disregarded. To be able to use the raw data, integration of the gyroscopes angle velocity is needed to express the data in angles. An offset error in the angle velocity for the Tobii glasses was found during analysis and must be regarded carefully if using the gyroscope data. This is due to small errors during integration will gradually increase and result in large errors.

Whether the cognitive task resulted in gaze concentration or not, has not been proven in this thesis. As a large difference in results was observed between participants for the cognitive task events and the previous baseline, the metric and eye tracking system that is used may influence how well gaze concentration is apparent in the results. Other types of cognitive tasks that would lead to higher levels of cognitive demand may have resulted in more events with obvious gaze concentration.

6 Conclusions

This thesis analyzed and applied seven of the most common gaze concentration metrics. The chosen metrics, PRC, PAI, SGE, TGE, SDRG, SDHG and SDVG, were applied to gaze data from both a 2-camera system and Tobii glasses while yielding slightly different results.

Analysis of the amount of missing data of the two eye tracking systems showed that the 2-camera system had 10% missing and invalid data compared to Tobii glasses that had 5% missing data. The amount of missing data can affect the results of the metrics when the calculations use a shorter time window.

The results from this thesis work showed that there were distinct differences between the 2-camera system and the Tobii glasses for PRC, PAI and TGE. However, a similar pattern was shown between the two systems for SGE, SDRG, SDHG, and SDVG. This demonstrates that comparing results between different research studies should be done with care unless the same eye tracking system is used.

PRC (percent road center) is a robust metric applicable to any type of gaze data if a road center area can be defined as it classifies gaze data as either within the road center area or not. Similarly, PAI (percent area of interest) is also easily applied to various types of gaze data but may not capture gaze concentration in all instances. This thesis work showed that an eye tracking system that enables a more precisely defined AOI for on-road glances is preferred for indicating gaze concentration.

Both SGE (stationary gaze entropy) and TGE (transitional gaze entropy) metrics require consideration of the type of gaze dispersion the metrics will be applied on. The size and number of bins used significantly impacted the entropy value, making it necessary to conduct a thorough analysis rather than applying the size and number of bins arbitrarily to any gaze data set. Both metrics were sensitive to outlying data and may produce misleading results when high uncertainty in gaze positions is present. Also, using a grid pattern of equally sized bins may not effectively capture gaze concentration in TGE, whereas an AOI-based definition of the bins could possibly yield more realistic results.

SDRG (standard deviation of radial gaze) and SDHG (standard deviation of horizontal gaze) both exhibit similar characteristics. These are robust metrics applicable to any type of gaze data. However, skewed results may occur when applied to eye tracking systems where gaze points are not accurately positioned such as manually mapped peripheral gaze points for the Tobii glasses. SDVG (standard deviation of vertical gaze) was shown to be a less sensitive metric, consistent with previous literature, and it was thus concluded that this metric was not suitable to capture gaze concentration.

No conclusion can be drawn regarding which metric is the best for capturing gaze concentration, since the true gaze directions were unknown. However, PRC, PAI, SGE, SDRG and SDHG show similar trends which indicate that they all capture the same gaze behavior at some level. Due to the small size of the dataset, it is difficult to state which metric that are more sensitive capturing gaze concentration compared to the others.

7 References

Abbasi, J. A., Mullins, D., Ringelstein, N., Reilhac, P., Jones, E., & Glavin, M. (2022, July). An Analysis of Driver Gaze Behaviour at Roundabouts. *IEEE Transactions on Intelligent Transportation Systems, 23*(7), 8715-8724.

Ahlström, C., Kircher, K., & Kircher, A. (2009). Considerations when calculating percent road centre from eye movement data in driver distraction monitoring. *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 132-139. Retrieved from https://www.researchgate.net/publication/242092761_Considerations_when_calculating_percent_road_centre_from_eye_movement_data_in_driver_distraction_monitoring

Duchowski, A. (2007). Eye tracking techniques. *Eye tracking methodology: Theory and practice*, 51-59. Retrieved April 12, 2023, from https://link.springer.com/content/pdf/10.1007/978-1-84628-609-4_5.pdf

Fredriksson, A., & Wallin, J. (2020). *Mapping an Auditory Scene Using Eye Tracking Glasses.* Linköping: Linköping Universitet. Retrieved from https://liu.diva-portal.org/smash/get/diva2:1479381/FULLTEXT01.pdf

Gonzalez-Sanchez, J., Baydogan, M., Chavez-Echeagaray, M. E., K. Atkinson, R., & Burleson, W. (2017). Chapter 11 - Affect Measurement: A Roadmap Through Approaches, Technologies, and Data Analysis. *Emotions and Affect in Human Factors and Human-Computer Interaction*, 255-288. Retrieved from https://doi.org/10.1016/B978-0-12-801851-4.00011-2.

Hammel, K. R., Fisher, D. L., & Pradhan, A. K. (2002). Verbal and Spatial Loading Effects on Eye Movements in Driving Simulators: A Comparison to Real World Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46*(26), 2174–2178.

Hansen, D. W., & Ji, Q. (2010). In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(3), 478-500. doi:doi: 10.1109/TPAMI.2009.30

International Organisation for Standardization. (2020). Road vehicles — Measurement and analysis of driver visual behaviour with respect to transport information and control systems. *ISO standard No. 15007:2020*. Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:15007:ed-1:v1:en:sec:F.1

Kountouriotis, G. K., Spyridakos, P., Carsten, O. M., & Merat, N. (2016). Identifying cognitive distraction using steering wheel reversal rates. *Accident Analysis & Prevention, 96*, 39-45. Retrieved April 4, 2023, from https://www.sciencedirect.com/science/article/pii/S0001457516302615

Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., Perilli, F., Pires, A., . . . Villalobos, N. (2015). Gaze Transition Entropy. *ACM Transactions on Applied Perception, 13*, 1-20. doi:10.1145/2834121

Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-Based Statistical Analysis of Eye Movement Transitions. *Association for Computing Machinery*, 159-166.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Retinal and Eye Research, 25*(3), 296-324.

Land, M., & Horwood, J. (1995). Which parts of the road guide steering? *Nature, 377*(6547), 339-340.

Lehtonen, E., Lappi, O., Summala, H., & Kotkanen, H. (2013). Look-ahead fixations in curve driving. *Ergonomics, 56*(1), 34-44. Retrieved from https://doi.org/10.1080/00140139.2012.739205

Li, P., Markkula, G., Li, Y., & Merat, N. (2018). Is improved lane keeping during cognitive load caused by increased physical arousal or gaze concentration toward the road center? *Accident Analysis & Prevention, 117*, 65-74. Retrieved from https://doi.org/10.1016/j.aap.2018.03.034.

Liang, Y., & Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention, 42*(3), 881-890. doi:https://doi.org/10.1016/j.aap.2009.05.001

Liang, Y., Lee, J. D., & Reyes, M. L. (2007). Nonintrusive Detection of Driver Cognitive Distraction in Real Time Using Bayesian Networks. *Transportation Research Record, 2018*(1), 1-8. doi:https://journals.sagepub.com/doi/10.3141/2018-01

National Center for Statistics and Analysis. (2021). *Distracted Driving 2019.* Washington, DC: National Highway Traffic Safety Administration. Retrieved January 26, 2023, from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813111

Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J., & Goodman, M. (2006). An overview of the 100-Car naturalistic study and findings. *National Highway Traffic Safety Administration*. Retrieved from https://www-esv.nhtsa.dot.gov/Proceedings/19/05-0400-W.pdf

Niezgoda, M., Tarnowski, A., K. M., & Tomasz, K. (2015). Towards testing auditory–vocal interfaces and detecting distraction while driving: A comparison of eye-movement measures in the assessment of cognitive workload. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 23-34.

Nilsson, E. J., Victor, T., Ljung Aust, M., Svanberg, B., Lindén, P., & Gustavsson, P. (2020). On-to-off-path gaze shift cancellations lead to gaze concentration in cognitively loaded car drivers: A simulator study exploring gaze patterns in relation to a cognitive task and the traffic environment. *Transportation Research Part F: Traffic Psychology and Behaviour, 75*, 1-15. Retrieved from https://doi.org/10.1016/j.trf.2020.09.013.

Pillai, P., Balasingam, B., Kim, Y. H., Lee, C., & Biondi, F. (2022). Eye-Gaze Metrics for Cognitive Load Detection on a Driving Simulator. *IEEE/ASME Transactions on Mechatronics, 27*(4), 2134-2141.

Radlmayr, J., Fischer, F. M., & Bengler, K. (2019). The Influence of Non-driving Related Tasks on Driver Availability in the Context of Conditionally Automated Driving. *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). 823*, pp. 295-304. Florence: Springer, Cham. doi:https://doi.org/10.1007/978-3-319-96074-6_32

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2012). A Field Study on the Impact of Variations in Short-Term Memory Demands on Drivers' Visual Attention and Driving Performance Across Three Age Groups. *Humans Factors, 54*(3), 454-468. Retrieved from https://doi.org/10.1177/0018720812437274

Schieber, F., & Gilland, J. (2008). Visual Entropy Metric Reveals Differences in Drivers' Eye Gaze Complexity across Variations in Age and Subsidiary Task Load. *Proceedings of the Human Factors and Ergonomics Society Annual*

Meeting, 52(23), 1883-1887. Retrieved from https://doi.org/10.1177/154193120805202311

Shannon, c. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, 27*(3), 379-423.

Shiferaw, B. A., Crewther, D. P., & Downey, L. A. (2019, November 1). Gaze entropy measures detect alcohol-induced driver impairment. *Drug and Alcohol Dependence, 204*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0376871619302789?via%3Dihub

Shiferaw, B. A., Downey, L. A., Westlake, J., Stevens, B., Rajatman, S. M., Berlowits, D. J., . . . Howard, M. E. (2018). Stationary gaze entropy predicts lane departure events in sleep deprived drivers. *Scientific Reports, 8*(220). doi:https://doi.org/10.1038/s41598-018-20588-7

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience and Biobehavioral Reviews, 96*, 353-366.

Tivesten, E., Broo, V., & Ljung Aust, M. (2023, May). The influence of alcohol and automation on drivers' visual behavior during test track driving. *Transportation Research Part F: Traffic Psycology and Behaviour, 95*, 215-227. doi:https://doi.org/10.1016/j.trf.2023.04.008

Tivesten, E., Victor, T., Gustavsson, P., Johansson, J., & Ljung Aust, M. (2019). Out-of-the-loop crash prediction: The Automation Expectation Mismatch (AEM) algorithm. *IET Intelligent Transport Systems, 13*(8), 1231-1240. Retrieved from https://doi.org/10.1049/iet-its.2018.5555

Tobii. (2022, November). *Tobii Pro Lab User Manual.* Retrieved April 4, 2023, from Tobii Connect: https://connect.tobii.com/s/lab-downloads?language=en_US

Tobii. (2023, January). Tobii. *Tobii Pro Glasses 3 User Manual(1.16)*. Retrieved April 4, 2023, from Tobii Connect: https://connect.tobii.com/s/g3-downloads?language=en_US

Tobii AB. (2023, April 14). Glasses 3 wearable eye tracker. *Press kit and media assets*. Retrieved April 14, 2023, from https://corporate.tobii.com/newsroom/press-kit-media-assets

Trafikverket. (2023, feb 3). *Rattfylleri*. Retrieved from trafikverket.se: https://www.trafikverket.se/resa-och-trafik/trafiksakerhet/sakerhet-pa-vag/rattfylleri/

Victor, T. (2005). *Keeping Eye and Mind on the Road.* Uppsala: Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 9. Retrieved Febuary 15, 2023, from https://www.diva-portal.org/smash/get/diva2:167500/FULLTEXT01.pdf

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*(2), 167-190. Retrieved from https://www.sciencedirect.com/science/article/pii/S1369847805000161

Victor, T., & Larsson, P. (2004). Method and arrangement for interpreting a subjects head and eye activity. *(Patent No. International publication number PCT WO 2004/034905 A1)*. Retrieved May 10, 2023, from https://patents.google.com/patent/US7460940B2/en

Victor, T., Dozza, M., Bärgman, J., Boda, C.-N., Engström, J., Flannagan, C., . . . Markkula, G. (2015). Analysis of Naturalistic Driving Study Data: Safer.

Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk, 10.17226/22297.

Volvo Cars . (2022, Sep 22). *Your Volvo car will understand you when you're not at your best – and step in when you need support.* Retrieved from [Press release]: https://www.media.volvocars.com/global/en-gb/media/pressreleases/304020/your-volvo-car-will-understand-you-when-youre-not-at-your-best-and-step-in-when-you-need-support

Wang, Y., Reimer, B., Dobres, J., & Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour, 26*(PA), 227-237. Retrieved from https://www.sciencedirect.com/science/article/pii/S136984781400120X

Appendix

A Environment conditions during the test study

Table 5: Conditions in the environment during the drive for the five participants.

	Test participant 1			Test participant 2		
Event	**Weather**	**Road conditions**	**Traffic environment**	**Weather**	**Road conditions**	**Traffic environment**
Baseline 1	Cloudy, light snow fall	Wet with snow on the sides of the road	One overtake by ego vehicle at start of event, two overtaking PC	Cloudy, moderate rain	Wet	Three overtaking PC, two overtaking buses in the right side bus-lane. Moderate traffic
Task 1	Cloudy, light snow fall	Wet with snow on the sides of the road	Two overtaking PC	Cloudy, moderate rain	Wet	Two overtaking PC
Baseline 2	Cloudy, light rain	Wet with snow on the sides of the road	One overtaking PC Two overtaking PC Lead vehicle: double decker bus approximatly 100m ahead	Cloudy, moderate rain	Wet	-
Task 2	Cloudy, light rain	Wet with snow on the sides of the road	Long overtake at start of event Lead vehicle: double decker bus approximately 100m ahead	Cloudy, moderate rain	Wet	Two overtaking PC
Baseline 3	Cloudy, light rain	Wet with snow on the sides of the road	Three overtaking PC	Cloudy, moderate rain	Wet	One overtaking PC
Baseline 4	Cloudy, light rain	Wet with snow on the sides of the road	Four overtaking PC	Cloudy, moderate rain	Wet	Four overtaking PC
Task 3	Cloudy, light snow fall	Wet with snow on the sides of the road	Four overtaking PC	Cloudy, moderate rain	Wet	Two overtaking PC
Baseline 5	Cloudy, moderate rain/snow	Wet with snow on the sides of the road	Two overtaking PC, moderate amount of traffic	Cloudy, moderate rain	Wet	Two overtaking PC
Task 4	Cloudy, light snow fall	Wet with snow on the sides of the road	Three overtaking PC	Cloudy, moderate rain	Wet	Four overtaking PC, one overtaking bus in the right side bus-lane. A truck driving closely behind ego vehicle. Five overtaking PC, one overtaking truck. Lead vehicle: truck approcimatly 60m ahead
Baseline 6	Cloudy, light rain	Wet with snow on the sides of the road	Three overtaking PC	Cloudy, moderate rain	Wet	

	Test participant 3			Test participant 4			Test participant 5	
Event	**Weather**	**Road conditions**	**Traffic environment**	**Weather**	**Road conditions**	**Traffic environment**	**Weather**	**Road conditions**	**Traffic environment**
Baseline 1	Cloudy with slight light at the horizon	Wet/drying	Two overtaking PC	Sunny with small distributed clouds	Dry	Two overtaking PC	Cloudy with slight light shining through	Dry	Two overtaking PC, a minibus is placed in the right side bus lane, moderate amount of traffic
Task 1	Cloudy with slight light shining through	Wet/drying	Three overtaking PC, one overtaking PC in the right side bus-lane One overtake by ego vehicle of PC travelling in the righ exit lane	Sunny with small distributed clouds	Dry	Two overtaking PC, one PC entering highway from right side lane	Cloudy with slight light shining through	Dry	One overtaking PC with trailer, a minbus entering highway from the right side lane
Baseline 2	Cloudy with slight light shining through	Wet/drying		Sunny with small distributed clouds	Dry	Long overtaking by ego vehicle, passing two trucks	Cloudy with slight light shining through	Dry	Two overtaking PC, one PC entering highway from right side lane
Task 2	Cloudy with slight light shining through	Wet/drying	Ten overtaking PC	Sunny with small distributed clouds	Dry	One overtaking PC. A truck driving closely behind ego vehicle.	Cloudy with slight light shining through	Dry	Eight overtaking PC Leading vehicle: Truck approximatly 50m ahead
Baseline 3	Cloudy with slight light shining through	Wet/drying	One overtake by ego vehicle, one overtaking PC	Sunny with small distributed clouds	Dry	Ongoing overtaking by ego vehicle in the start of the event. Overtaking two trucks	Cloudy with slight light shining through	Dry	Four overtaking PC. Ongoing overtaking by ego vehicle in the start of the event. Overtaking one truck.
Baseline 4	Cloudy with slight light shining through	Wet/drying	One overtaking PC	Sunny	Dry	One overtaking PC	Cloudy with slight light shining through	Dry	Three overtaking PC
Task 3	Cloudy with slight light shining through	Wet/drying	Two overtaking PC	Sun ahead, clear sky	Dry	One overtaking PC, several PC entering highway from right side lane	Cloudy with slight light shining through	Dry	One overtaking PC, one PC entering highway from right side lane
Baseline 5	Cloudy with slight light shining through	Wet/drying	Two overtaking PC	Sun changes from ahead to left side (visor is moved), clear sky	Dry	Three overtaking PC Leading vehicle: Truck approximatly 60m ahead	Cloudy with slight light shining through	Dry	One overtaking PC, one PC exiting the highway
Task 4	Cloudy with slight light shining through	Wet/drying	Four overtaking PC	Sun directed from the left, clear sky	Dry	Three overtaking PC Leading vehicle: Truck approximatly 60m ahead	Cloudy with slight light shining through	Dry	One overtaking PC one PC entering highway from right side lane
Baseline 6	Cloudy with slight light shining through	Wet/drying		Sun directed from the left, clear sky	Dry	One overtake by ego vehicle of truck	Cloudy with slight light shining through	Dry	Three overtaking PC

B Density plots with grid size

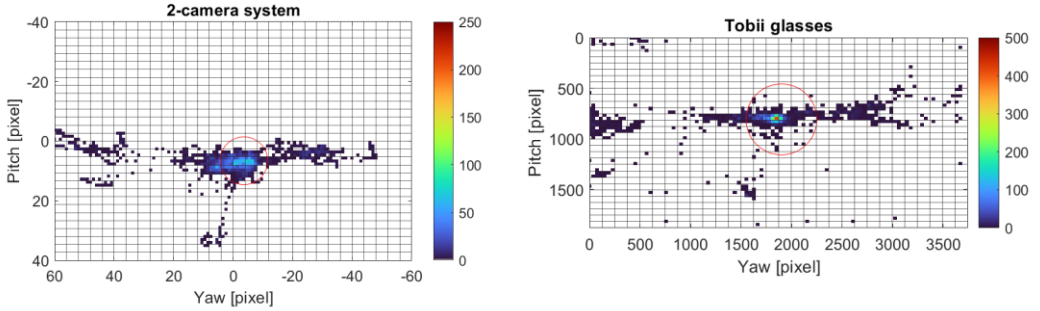


Figure 36: Illustration of grid size 30x30. Density plot for Participant 2 - Baseline 2.

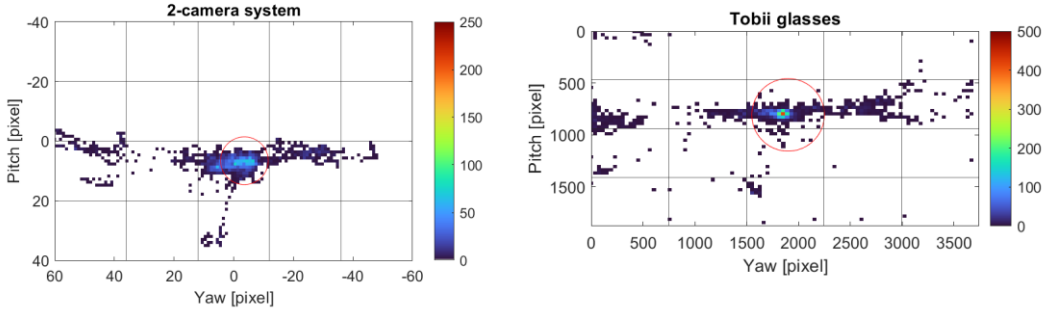


Figure 37: Illustration of grid size 4x5. Density plot for Participant 2 - Baseline 2.

DEPARTMENT OF MECHANICS AND
MARITIME SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY