Received: 15 June 2023 | Revised: 1 April 2024 | Accepted: 3 June 2024

IET Intelligent Transport Systems

DOI: 10.1049/itr2.12520

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Practical aspects of measuring camera-based indicators of alcohol intoxication in manual and automated driving

Raimondas Zemblys[1] | **Christer Ahlström**[2,3] | **Katja Kircher**[2] | **Svitlana Finér**[1]

[1]Smart Eye AB, Göteborg, Sweden

[2]Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden

[3]Department of Biomedical Engineering, Linköping University, Linköping, Sweden

Correspondence
Christer Ahlström, Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden.
Email: christer.ahlstrom@vti.se

Funding information
VINNOVA FFI, Grant/Award Number: 2019-05834

Abstract

Camera-based Driver Monitoring Systems (DMS) have the potential to exploit eye tracking correlates of alcohol intoxication to detect drunk driving. This study investigates how glance, blink, saccade, and fixation metrics are affected by alcohol, and whether possible effects remain stable across three different camera setups, as well as when the driver is out-of-the-loop during level 4 automated driving (Wizard-of-Oz setup). Thirty-five participants drove on a test track first sober and then with increasing intoxication levels reaching a breath alcohol concentration (BrAC) of 1‰. Linear Mixed-Effects Regression analyses showed that with increasing intoxication levels, eye blinks became longer and slower, glances and fixations became fewer and longer, and more attention was directed to the road area, at the expense of more peripheral areas. Fixation and blink metrics were more robust to changes in automation mode, whereas glance-based metrics were highly context dependent. Not all effects of alcohol intoxication could be measured with all eye tracking setups, where one-camera systems showed lower data availability and higher noise levels compared to a five-camera system. This means that lab findings based on higher quality eye tracking data might not be directly applied to production settings because of hardware limitations.

1 | INTRODUCTION

Drunk driving is a major road safety problem. Official statistics of drunk driving deaths are often under-reported, but it is estimated that around 25% of all road traffic fatalities in Europe are alcohol related [1]. In the U.S., alcohol-impaired driving fatalities accounted for 30% in 2020 [2, tab. 11]. Vissers et al. [3] concluded that among 45 countries worldwide a weighted average of 21.8% of road deaths were alcohol-related. Even low levels of alcohol lead to an increased risk of traffic crashes [1, tab. 3.1] and an increased risk of severe injuries and fatalities [4].

Driving consists of a complex interaction of situation anticipation, information sampling, decision making, and action, where the driver interacts with the environment, other road users, and the vehicle. The driving process is coordinated by complex interactions encompassing operational, tactical and strategic abilities. If functioning of one or more of these elements is degraded, the driver might become unfit to drive. Previous studies have shown that several driving impairments arise when driving under the influence of alcohol. Examples include decreased smoothness of motor functions which affect steering and lane keeping [5–7], increased impulsivity, leading to risk taking such as speeding and reduced likelihood to stop at red lights [8], or impaired cognitive functioning, which may lead to increased reaction times and compromised decision-making [9–11].

A wide range of policies and countermeasures are needed to reduce motor vehicle crashes and fatalities caused by alcohol. These countermeasures include reducing the legal alcohol concentration limit for driving, enforcing the alcohol limits, supplementing enforcement by public education campaigns, and introducing punishments with severe personal consequences (such as on-the-spot fines, driving licence penalty points and, as appropriate, driving licence suspension, and alcohol ignition locks). In addition, there is a growing interest for continuous intoxication monitoring. For example, EuroNCAP aims to include detection of driving under the influence in their ratings by 2030 [12], thus potentially making it mandatory in all passenger vehicles. Similarly, the National Transportation Safety Board in US is calling[1] for alcohol detection systems in all new vehicles, showing general interest in using advancements in technology to prevent drunk driving.

Several approaches for intoxication detection while driving have been proposed. These include contact and remote breath analysis [13, 14], skin conductivity analysis [15], heart rate variability [16], gaze behaviour and head movements [17], driving behaviour [18], or a combination of the above. Most of the methods require additional and sometimes obtrusive sensors to be installed in the car, or are not feasible in case of automated driving. Driver monitoring cameras are already on the verge of being installed in every newly produced vehicle, making it a very attractive approach for estimating driver state, including various types of intoxication. Moreover, these systems can be used to monitor drivers under different levels of vehicle automation thus enabling estimation of drivers' fitness to take over.

Camera-based eye tracking and eye movement measures in general have proven to be useful tools for investigating alcohol intoxication impairments and have been extensively studied before [19]. For example, alcohol intoxication has been found to reduce smooth pursuit gain, leading to higher rates of catch-up saccades and impaired perception of depth from motion parallax [20], to reduce saccade latency, accuracy and velocity [21, 22], to increase attention bias [23, 24], to reduce saccadic distance and number of fixations [25], as well as to decrease the correlation between eye and steering movements, even at low intoxication levels [26]. However, it remains unclear if findings from lab and simulator studies are valid also in real driving settings with production grade sensors.

The aim of this paper is to evaluate how various driving performance related behavioural and psychophysiological eye tracking measures are affected by alcohol intoxication and whether these effects can be reliably estimated under realistic driving conditions with naturalistic illumination and actual vehicle and driver movements. We also analyse which measures are robust enough and can measure effects of intoxication if using a single-camera Driver Monitoring System (DMS) mounted on the steering wheel column (most common DMS placement) as well as when the driver is out-of-the-loop during highly automated driving. Development of an actual intoxication detection algorithm is out of the scope of this paper, however, the results presented here will bring us one step closer to a reliable, driving mode agnostic, remote intoxication detection system.

The main contribution of this paper is to provide evidence that both high-level behavioral and lower-level psychophysiological eye tracking measures are affected by alcohol intoxication in both manual and automated driving. We also show that lower-level psychophysiological measures reflect the driver's state better as the effect of intoxication is less dependent on the driving mode. Finally, we demonstrate the importance of taking eye tracking data quality into account when interpreting the results from different sensor configurations.

[1] https://www.ntsb.gov/news/press-releases/Pages/NR20220920.aspx, accessed April 5, 2023

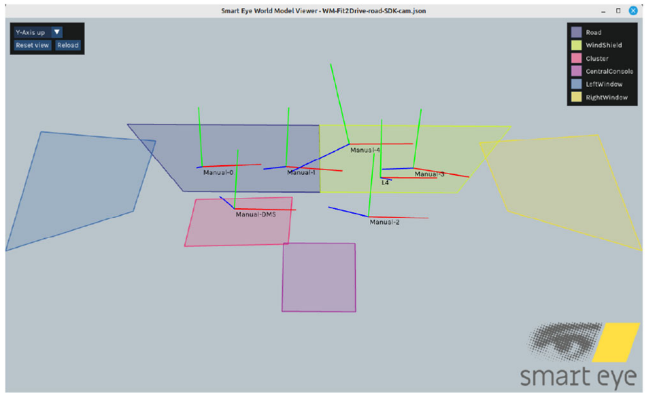

FIGURE 1 Test vehicle world model, showing areas-of-interest (AOIs) together with camera positions for the *Manual*, *Manual-DMS* and *L4* setups. Red, green and blue lines denote x, y, and z axes of the cameras. In the automated drives, the *Road* and the *WindShield* AOIs switched places.

2 | METHOD

In this study, participants performed controlled test track drives, first completely sober as control and then under the influence of alcohol (around 0.2, 0.5, 0.8 and 1.0‰ Breath Alcohol Concentration – BrAC). The test track drives, consisting of manual and Level 4 automated[2] drives in a Wizard-of-Oz setup, captured data from a close-to-naturalistic driving setting, except that no other road users were present for safety reasons. At the same occasion, a driving simulator study was performed [27]. Simulator data were not used in the present paper.

During the manual test track drives, two separate camera setups were used to record the participants. One system comprised five dashboard cameras, and the other system was a one-camera system placed on the steering column (a typical location for a DMS camera; see Figure 1). These system setups are referred to as the *Manual* setup and the *Manual-DMS* setup, respectively. The five-camera setup was meant to represent an idealized research setup, to ensure that driver behavior was captured from multiple angles, thus avoiding data loss because of, for example, occlusions. The *Manual* setup thus serves as a reference to which other setups are compared. The single-camera *Manual-DMS* setup on the other hand represents a realistic hardware configuration that could be expected in a passenger vehicle. In the automated level 4 driving sessions, a slightly different one-camera system mounted on the dashboard was used. This eye tracking system is referred to as the *L4* setup and was also meant to represent a camera setup applicable in a real world environment.

2.1 | Participants

The participants were recruited by word of mouth and a list of interested participants. Screening was done via a recruitment questionnaire that encompassed driving background, drinking

[2] https://www.sae.org/standards/content/j3016_202104/

habits and relevant physiological information. Incompatible medical conditions or pregnancy belonged to the exclusion criteria, likewise self-reported aggressive behaviour under the influence. Participants were recruited in pairs as far as possible, preferably friends or couples. The reasons for this were both of practical and social nature (it was assumed that the experimental situation would be more relaxed for the participants if they were in the company of someone they knew). A total of 35 participants were recruited for the experiment (15 females, ages 26–60, mean 43.7 years, 20 males, ages 22–60, mean 38.3 years) and completed the task. Each participant received 1000 SEK before tax as compensation for their time, effort and travel costs.

2.2 | Test procedure

Recording sessions took place on a closed racing track and its facilities in Mantorp Park, Sweden on weeks 43 and 45, 2021. Several days before the experiment, the participants received written information about the study, the planned procedure and practical issues, such as instructions to have a good night's sleep before the study, to get a healthy breakfast etc., per e-mail. The participants also received a consent form that was to be signed by a person able to pick up the participant(s) after the trials, guaranteeing that they would take care of the intoxicated person(s). Upon arrival at the test site, participants received an oral recapitulation of the procedures and were shown all relevant facilities. Female participants were offered to take a pregnancy test. The participants signed an informed consent form and a form for monetary reimbursement as well as a background questionnaire. They also took a breath analyzer test to confirm to have a level of 0.0‰ BrAC. Participants were informed that if they or one of the test leaders decided that the participant was not capable of continuing the experiment, the data collection for this person was to be terminated without reduction of monetary compensation or any other consequences. None of the participants were terminated prematurely.

There were two recording sessions per day, a morning session from around 09:00–14:00 and an afternoon session from around 14:30–19:30. In each except one of the sessions two participants took part, where one in the pair was randomly assigned to start in the driving simulator [27], and the other on the test track. Then they switched so both participants took part of both the simulator and test track drives.

The test procedure was as follows:

1. the participant was seated in either the simulator or the driver seat in the car. If required, participant adjusted car mirrors, seat etc.
2. the test leader informed about the upcoming drive and the participant was asked to look at six *gaze calibration points* on the computer screens in the driving simulator, or in the car interior or. After that, they drove either part of the test route (ca. 5 min) in the simulator or one lap on the test track in manual mode to familiarize themselves with the simulator/car and the environment.
3. the participant drove the full simulator route or one lap on the test track in manual mode, followed by one lap in automated L4 mode, always in that order (further *Trial 1*).
4. the participant drove in simulator if they started with test track drives, and the other way around (further *Trial 2*).
5. the participant was then escorted to the test facility where they consumed alcohol and repeated all, except step 2, while intoxicated.

2.2.1 | Alcohol intake

The participants were offered vodka, white rum and gin, which could be mixed with a soft drink of choice. Participants had around 15 min to drink, followed by a 5-min period to let alcohol in the mouth to wear off, after which they rinsed their mouth with water. For the first alcohol intake the participants drank the amount of 1.5-2 standard glasses, depending on body weight and gender according to Hume et al. [28]. The aim was to achieve around 0.2‰ BrAC after 20 min. The BrAC target levels used in the study were 0.2, 0.5, 0.8, and 1.0‰. Adjustments to alcohol amount were made to later drinking conditions if the measured BrAC level was above or below the target level.

In between driving, participants were offered snacks like chips, nuts and fruit as well as the possibility to use the bathroom. To avoid potential problems with how food consumption may influence the alcohol intoxication level of the participants, the intention was to not offer any large meals during the course of the experiment. However, should a participant feel unwell or drained of energy, there was a possibility to get a sandwich or similar, which a few participants made use of.

2.2.2 | Test track drives

The Mantorp test track is designed for racing, therefore the road is wide with run-off areas next to the track. The section of the track used in the study was about 3 km long, was driven in clockwise direction and took around 4 min to complete. In places where there was an uncertainty of which direction to take, cones indicated the correct path. To simulate realistic driving conditions, the track was equipped with several speed limit signs (50 and 70 km/h), a narrow passage (4 m wide and 30 m long), a chicane (designed after [29] Table 1, Scheme No. 20), as well as warning signs 25 m before the narrow passage and the chicane, and a "Give way" sign 75 m before the start/stop point of the track (see Figure A1 in Appendix A).

Both manual and level 4 automated driving were tested. Manual driving was operationalised as standard driving, where the participant sat in the driver's seat and drove without speed or lane keeping assistance. Due to the nature of the test track (no standard lane markings, wide road), lane keeping assistance did not function, why the automated mode was simulated such that the participant sat in the passenger seat, with the test leader driving around the track. The participants were told to pretend that they were driving a car with a reliable Level 4 automation system

TABLE 1 Eye tracking system specifications.

Setup	Camera	Number of cameras	Resolution	Placement
Manual	Basler acA1300-75gm	5	1280 × 1024	Dashboard and interior, see “Manual–0 through 4” in Figure 1
Manual-DMS	AI-X	1	1600 × 1300	Steering column, see “Manual-DMS” in Figure 1
L4	Blackbird3P	1	1280 × 752	Dashboard on the passenger side, see “L4” in Figure 1

(the functionality of which was explained to the participant) and to behave as they saw fit with a system of that kind.

The test vehicle was a passenger car (2013 Volkswagen Passat, automatic gearbox, left-hand drive) with an additional set of accelerator and brake pedals on the passenger side. During the simulated automated drives, the auxiliary pedals were covered to avoid intoxicated participants accidentally interfering with vehicle operation. Prior to the test sessions, test leaders underwent a safety driving training program and were trained to take over in the case of an emergency.

2.3 | Data recording equipment and measurements

Participants' BrAC was measured with a *Dräger Alcotest 6820* breathalyzer. BrAC measurements were taken before the manual and after the automated test track drive, and before and after the simulator trial. In addition, BrAC was measured occasionally after consuming alcohol to see when the BrAC level stabilised. Two breathalyzers, both of which were calibrated before the recordings, were randomly used for measurements without any particular preference.

The test vehicle had three eye tracking systems which were all synced and recorded at 60 Hz sampling rate (see Table 1). In addition, one camera recorded the forward scene, while a wide field-of-view RGB-IR camera recorded the car interior (the latter was not used in this study). All recordings were stored on an external hard disk drive for offline analysis. The test vehicle was also equipped with a CAN logger that recorded vehicle speed, throttle, brake etc. signals and a GPS tracking unit for vehicle position estimation. Both loggers were synced with the camera recordings. Since camera and vehicle data recordings started before onset of the test trial, vehicle speed and GPS data were used to manually confirm the exact start and end times of each trial by looking at the scene recording from one of the cameras for a frame where the car passed the start line of the track or stopped after completing a lap.

The participants rated their sleepiness level on the *Karolinska Sleepiness Scale* [30] before and after each simulator and test track drive. Before each drive, participants rated how good (on a scale from 0 to 10) they expected to drive, while afterwards they rated how well they had actually driven. After each drive, the test leader annotated the participant's mood based on an observation of the participant's general behaviour (using a set of descriptors: “relaxed”, “nervous”, “stressed”, “communicative”, “tense”, “sleepy”, and “nausea”). An electrocardiogram (lead II) was also recorded using a *TEMEC Vitaport 2* bio-amplifier, but the data were not used in this study.

2.3.1 | Pre-processing of gaze data

Gaze direction[3] and eyelid opening were extracted from the camera recordings using *SmartEye Pro 10.1.3*. Gaze data for each of the drives were then pre-processed offline by first removing short (≤100 ms) sequences of data that had data loss episodes before and after, followed by linear interpolation of short (≤50 ms) data loss episodes. This ensured that short data loss episodes did not fragment, for example, ongoing fixations or saccades, while at the same time guaranteed that only data that spanned longer than short fixations were considered.[4] Note that it is common with eye tracking data loss, for example during eye blinks or when the cameras' view of the participant's face is obstructed by the hands, the steering wheel or other objects. It is therefore common to interpolate or in other ways handle data loss before further gaze data analyses [see e.g. [32, 36–38]].

Gaze data with quality above 0.0 as reported by the *Smart Eye Pro* software were considered to be valid. Data loss episodes with duration ≤200 ms and data during blinks were filled-in with nearest sample interpolation with a small additive Gaussian noise.[5] While this is not an accurate representation of real gaze data, it is sufficient for the purposes of subsequent eye movement event detection, by ensuring that ongoing fixations are not fragmented due to data loss or blinks. If the gaze direction before and after the blink/data loss episode differ, the interpolation procedure simulates a saccade with a plausible amplitude (albeit with inaccurate duration and peak velocity, and potential lengthening of adjacent fixations). Only blinks with at least one invalid gaze data sample, valid samples immediately before or after the blink, and duration less or equal to the *extreme outlier*[6] threshold were considered for the fill-in procedure. The blink duration threshold was calculated independently for each drive, effectively adapting to individual differences and whatever effects the intoxication level may have on the blink duration. Such adaptive thresholding ensures that the data loss fill-in

[3] The *estimated gaze* signal expressed as *heading* and *pitch* was used here [31]

[4] See works by, for example, Zemblys et al. ([32] Figure 1 and [33] Figure 3) who showed that the shortest fixations are just below 100 ms, although others report fixations as short as 30–50 ms or even less ([34] Figure 11.14; [35], Table 2)

[5] Variation in the gaze signal – here $N(\mu = 0, \sigma^2 = 0.01°)$ – is only required for numerical stability. The IRF eye movement event detector [32] used here requires calculation of the standard deviation of the data window.

[6] Blinks that last longer than $Q3 + 3 \times IQR$, that is, longer than 3 times the interquartile range (IQR) from the upper quartile $Q3$ [39].

procedure does not bias subsequent analyses, by only filling in blinks typical for a particular driver in a particular condition.

Eye movement event detection and glance calculation

The Identification by Random Forest (IRF) event detector [32] was used for fixation and saccade detection. Originally, IRF was designed to detect three event types, fixations, saccades and post-saccadic oscillations, but here, IRF was retrained to only detect fixations and saccades.[7] All the original IRF features except *sampling frequency (fs)*, *Rayleigh test*, and *i2mc* were used [32, Table 1]. Also, fixation merging post-processing settings were adjusted to merge fixations which were less than 400 ms and 2 degrees apart (instead of the original 75 ms and 0.5 degrees) since detected saccade durations were allowed to be up to 400 ms. Similarly to the procedure described in [32], the remaining gaze data loss episodes were interpolated using the same nearest sample interpolation approach as previously described, and later events corresponding to the interpolated samples were set to *undefined*. Fixations and saccades adjacent to the *undefined* events, together with the first and the last event in each of the drives, were removed since they are likely only partial events and might skew eye movement event measures or affect further analyses.

Following eye movement event detection, *glances* were calculated. Glances are here defined as events that start with the onset of the first fixation on an area-of-interest (AOI[8]) and end with the offset of the last fixation on that same AOI. A total of six AOIs were used in the study: *Road*, *WindShield*, *Cluster*, *CentralConsole*, *LeftWindow*, and *RightWindow* (see Figure 1). Together, the *Road* and *WindShield* AOIs represent the physical windshield, which in the world model is split into two separate areas. Glances through the left half of the physical windshield were assumed to represent attention to the forward road, while glances through the right half were assumed to represent attention to the adjacent driving environment, such as road signs or sharp right curves. In the *L4* setup, the *Road* and the *WindShield* AOIs swapped places to account for the participant sitting in the passenger seat, and therefore glances to the *WindShield* represent left-hand glances to the driving environment. Size and position of each AOI were determined using a *Laser Chessboard Tool* [31]. To account for gaze inaccuracy, the size of smaller AOIs (*Cluster* and *CentralConsole*) were enlarged by adding a 5 cm border.

Before making any glance calculations, nearest sample interpolation was used to interpolate AOI data, to ensure that all interpolated gaze data had a corresponding most probable gaze intersection with an AOI. Each fixation event was then associated with the most frequent AOI value within that fixation. *Undefined* eye movement events together with discarded adjacent fixations and saccades were labelled as *Unknown* glance events, which were later discarded from the analyses if they were $\leq$1 s. This procedure ensured that only prolonged *Unknown* glance events were considered as shifts of attention. For example, if a short *Unknown* glance event was located between two glances towards the same AOI, then the two adjacent AOI glances were merged. Remaining *Unknown* glance events were interpreted as cases where it is impossible to determine towards what/where the driver was looking.

2.3.2 | Statistical analysis

Statistical analysis was performed using Linear Mixed-Effects Regression (LMER) modeling which was implemented in *R* version 4.3.0 [40] using the *lme4* [41] and the *lmerTest* [42] packages. P values were obtained using Satterthwaite‘s approximation. In addition to the marginal and conditional R^2 for model fits, the *Intraclass Correlation Coefficient* (ICC) will be reported as a measure of the proportion of the variance explained by the grouping structure. The exact model details are provided below in relevant sections.

3 | RESULTS

First, the measured BrAC level and gaze data quality were analysed (Sections 3.1 and 3.2). This was followed by an investigation of the effects of intoxication on behavioural and psychophysiological measures (Sections 3.3 and 3.4).

3.1 | Intoxication level

643 out of 700 BrAC measurements were available for further analysis.[9] On average, the measured BrAC corresponded well to the targeted level, although there were some between- and within-participant variation (see Figure 2). A few obvious outliers were present in the data, therefore we removed them using *Tukey's fence*[10] method applied within each intoxication condition and measurement order. A total of 13 outliers were removed.

LMER analysis showed that, except for the 0.2‰ intoxication condition, the BrAc level increased significantly by 0.04–0.07‰ when measured after performing *Trial1* (see Table B1 in Appendix B). This indicates that the peak BrAC level had not been reached before the driving task started. In the 0.2‰ intoxication level condition, BrAC decreased by 0.04‰ ($p = 8.92e-03$) after *Trial1* and remained the same throughout *Trial2*. In other intoxication conditions, BrAC decreased by 0.04–0.06‰ (although not statistically significantly in the 0.8‰ condition) after *Trial1* and did not change significantly throughout *Trial 2*.

Based on modeling results for each trial, a *ground-truth* BrAC was defined as follows: missing or removed outlier data were replaced by LMER model predictions, and measurements

[7] Post-saccadic oscillation samples were relabeled to be part of the saccade.

[8] Gaze point (and consequently a fixation) is considered to be "on AOI" when the estimated gaze direction intersects with an object defined in the world model. In case of intersections with multiple objects, the closest one to the viewer is considered to be looked *at* (or *through* in case of an object being, for example, car windows).

[9] 35[*participants*] × 4[*measurements*] × 5[*intoxication levels*] = 700.

[10] Outliers are defined as values falling outside 1.5 times the interquartile range from the quartiles [39].

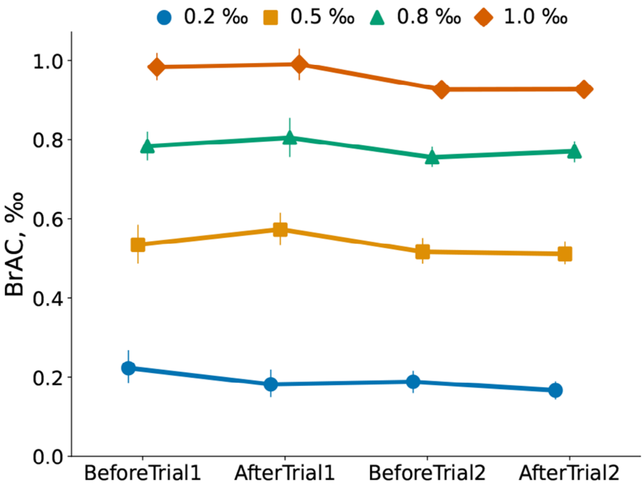

FIGURE 2 Average BrAC for all participants throughout the recording sessions. *Trial1* and *Trial2* refers to the chronological order of the drives, where half of the participants started to drive on the test track and the other half started in the simulator. Error bars show 95% confidence interval.

TABLE 2 Eye tracking data quality. μ and σ are averages and standard deviations across the participants. Note that the *estimated gaze* signal from *Smart Eye Pro* was used, which does not depend on corneal reflections and does not need gaze calibration [31] (Section 2.5).

Setup	Data loss (%) μ	Data loss (%) σ	Accuracy (degrees) μ (# of drivers)	Accuracy (degrees) σ	RMS (degrees) μ	RMS (degrees) σ	STD (degrees) μ	STD (degrees) σ
Manual	0.46	0.52	4.28 (35)	1.49	1.18	0.24	1.21	0.25
Manual-DMS	13.25	8.05	4.05 (30)	1.24	2.0	0.31	1.68	0.28
L4	3.04	4.32	-	-	1.83	0.29	1.67	0.29

before and after the trial were averaged. These averages were then used in the subsequent analyses.

3.2 | Gaze data quality

Out of 525 recordings, 12 were missing because of test leader mistakes or recording equipment failures. In addition to that, 16 recordings were removed due to high gaze data loss (more than 50% after gaze pre-processing), due to only partially recorded drives (shorter than 2 min), or due to recordings where less than three intoxication levels remained available for the driver in a certain recording setup. Reported results therefore include data from 497 recordings, yielding a recording attrition rate of 5.33%. All recordings from the *Manual* setup were available, whereas recordings from one to five drives on various intoxication levels were missing in the *Manual-DMS* and *L4* setups.

Gaze data loss was evaluated using post-processed gaze data. On average, gaze data loss from the five-camera *Manual* setup was 0.46% (see Table 2). The amount of data loss remained stable across the intoxication conditions (see left of Figure C1 in Appendix C). With the one-camera *Manual-DMS* and *L4* setups, the average amount of data loss increased from 11.65% to 15.3% and from 1.5% to 5.1%, respectively, in sober drives compared to the 1.0‰ intoxication condition drives. This suggests that with increasing intoxication level, the driver behaviour changed in a way that negatively affected data quality from the single-camera systems.

A calibration procedure was used where the participants looked at six calibration points in the car interior. The calibration procedure was conducted before the first manual drive, why accuracy can only be evaluated for the *Manual* and *Manual-DMS* recordings in the sober condition. The participants used a physical button to indicate when they were looking at each of the six calibration points. Accuracy was calculated as the absolute angular error between the estimated gaze direction and a *ground-truth* gaze direction (the vector pointing from the estimated gaze origin to the known 3D position of the calibration point), based on 1 seconds of data starting with the psychical button press. The participant-level mean and standard deviation of the median accuracy across calibration points is provided in Table 2. The accuracy in *Manual* setup was 4.3 degrees, while in the *Manual-DMS* the accuracy was 4.05 degrees (note that only 30 out of 35 participants had calibration data available in the *Manual-DMS* setup).

Gaze precision was estimated using a method adapted from [43] as the root mean square of the difference between successive gaze position samples (RMS) and the standard deviation of samples (STD). First, precision measures were calculated using a sliding window of 30 samples (0.5 s) in each of the detected fixations longer than the window size. Then for each of the recordings, all windows that had previously interpolated data were excluded before calculating the median value. Averages of the median RMS and STD, per driving mode, are provided in Table 2. Data in the *Manual* setup had an RMS of around 1.2 degrees and remained stable across intoxication conditions. The *Manual-DMS* and *L4* setups contained more noise, and the *L4* recordings showed a tendency to be noisier at higher intoxication levels. This again indicates that intoxication affect driver behaviour in a way that leads to reduced data quality in one-camera setups, which in turn may affect eye tracking metrics of alcohol intoxication.

3.3 | Behavioural measures

A glance analysis was conducted to assess how drivers' gaze behaviour changed when they were intoxicated. Glances provides information about where, when and for how long participants look, and are therefore often used as a surrogate measure of attention [44, 45, to name a few]. Various glance metrics are here considered to be higher-level behavioural indicators of how drivers perceive and interact with the world. In contrast, lower-level psychophysiological measures (described in Section 3.4) mostly reflect properties of oculomotor and other brain functions, that are usually not consciously controlled. It is expected that contextual factors (here reflected in driving modes and the camera setup used) will affect higher-level

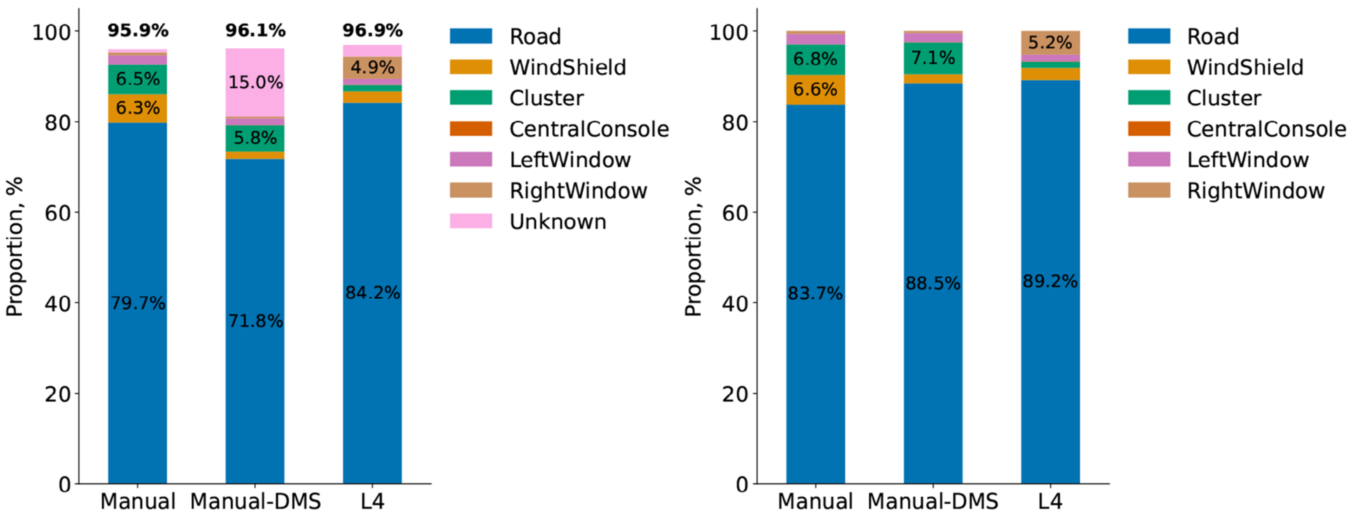

FIGURE 3 Total glance duration proportion for all AOIs in sober condition. Left – proportion relative to the total duration of the trials, right – normalized proportion, excluding *Unknown* glance events. Note that statistics here are different from the ones in Figures 6 and 8. Reported proportions reflect the *total* time that *all* participants glanced at each AOI.

glance metrics to a larger extent than they will affect lower-level metrics.

3.3.1 | Glance distribution

Figure 3 shows the total proportion of time that drivers glanced at different AOIs *in the sober condition*. Figure 3 (left) illustrates the proportion of glances relative to the total duration of the trials, including long (>1 second) *Unknown* glance events where gaze could not be tracked or eye tracking data was discarded in the pre-processing procedure. Note how proportions do not add up to 100%. In different recording setups, 3.1–4.1% of the data were not included in the glance data. These include the first and the last gaze events of each of the recordings that were discarded during preprocessing, together with saccades and short *Unknown* glance events in-between the different AOIs. The total proportion measure shows the amount of data that is actually available for further analyses. While the proportion of unaccounted data in the *Manual* and *Manual-DMS* setups was nearly identical, the proportion of *Unknown* glance events was considerably higher in the *Manual-DMS* (15% compared to 0.7%). These results reveal the limitations of the *Manual-DMS* setup, where data loss resulting from occlusions by the hands and the steering wheel affects the glance distribution. In the *L4* setup, recordings were made with one camera placed on the dashboard. In this setup, the camera view is less likely to be occluded, but extreme gaze angles may still result in gaze data loss. Here the proportion of *Unknown* glance events was 2.5%. This shows that the number of cameras and camera placement has a direct impact on the amount of usable data, which in turn affects further analyses and the conclusions that can be drawn.

Figure 3 (right) shows the normalized proportion of time that drivers were glancing at different AOIs (only including data where the gaze target could be determined). The drivers glanced at the road for 84-89% of the time in all setups. With the *Manual* setup, glances to the *WindShield* and the *Cluster* amounted to 6.6% and 6.8%, respectively. In the *Manual-DMS*, many of the glances to the *WindShield* were missed, leading to an apparent increase in the proportion of glances to the *Road* and to the *Cluster*. Note that the normalized glance proportion measure depends greatly on data loss, which in turn depends on the camera-systems used. For example, while the normalised glance proportion to the *Cluster* appears to be the same in the *Manual* and *Manual-DMS* setups, the proportion to the *Road* and the *WindShield* differs substantially (5% more to the *Road* and almost 5% less to the *WindShield*, respectively). Considering the total glance proportion (Figure 3, left), it is likely that many glances to these areas could not be registered, which affects the normalised distributions in Figure 3 (right).

Compared to the manual drives, the participants spent more time glancing to the road in the automated drives. This may be because the sober drives were done first, when the drivers did not yet know what to expect when the car was driven by a test leader simulating L4 automated driving, and therefore concentrated on the road. In the following recording (i.e. 0.2‰ intoxication condition), drivers seemed to feel more accustomed and allowed themselves to look around more. This observation also aligns with driver mood annotations (see Figure 8), where some participants reported being "nervous" and "tense" in the first drive.

3.3.2 | Glance transitions

Glance proportions only indicate how much the drivers looked at different AOIs, but it does not say anything about how they transitioned between different AOIs. Transitions between the AOIs in the *Manual*, *Manual-DMS*, and *L4* setups in the sober condition are shown in Figure 4. In all setups, the vast majority

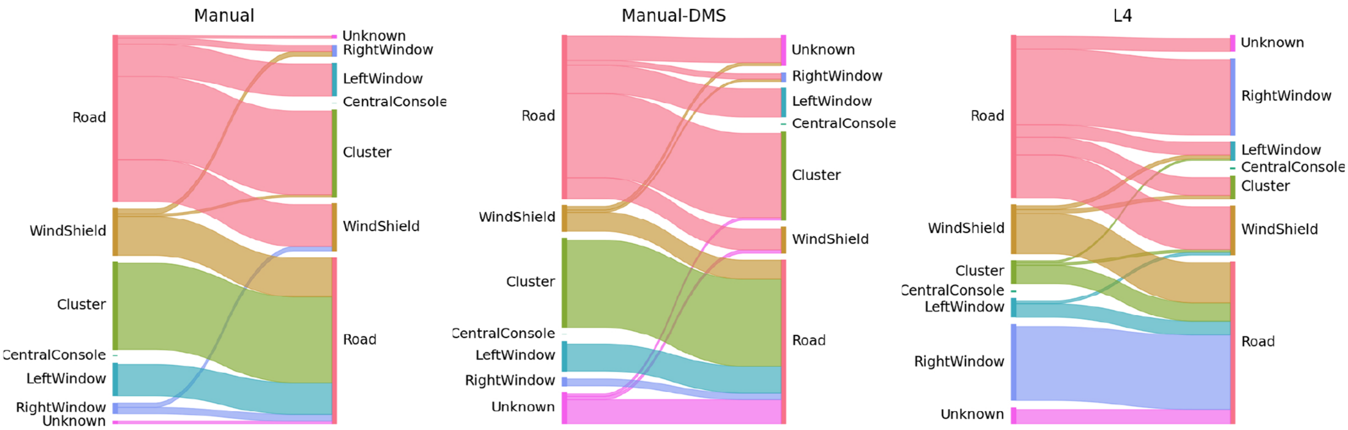

FIGURE 4 Sankey diagrams for the total number of transitions between the different AOIs. Left – *Manual*, middle – *Manual-DMS*, and right – *L4* setups, in the sober condition. Transitions with ≤0.5% occurrence are not shown.

of glances shifted to and from the *Road*, and only very few glances moved between two AOIs other than the *Road*.

Most of the glance transitions in the *Manual* setup (Figure 4, left) were between the *Road* and the *Cluster* (*Road* → *Cluster* – 24%, *Cluster* → *Road* – 25%), the *Road* and the *WindShield* (23%), and the *Road* and the *LeftWindow* (18%). Altogether almost 96% of the transitions were between the *Road* and other AOIs, which further confirms that drivers were very attentive to the road, while also monitoring their speed and ensuring situational awareness via glances to more peripheral areas of the driving environment. Similarly, in the *Manual-DMS* setup, 94% of the glance transitions were to and from the *Road*, with 49% of them being between the *Road* and the *Cluster*. Glance transitions between the *Road* and the *LeftWindow* constituted 16%, that is, despite of using only one camera, most of them were captured (cf. 18% in the *Manual* setup). However, the registered proportion of the *Road* ↔ *WindShield* transitions was considerably lower compared to when using the *Manual* setup (12% compared to 23%). Drivers were more likely to look through the *WindShield* when turning right on the test track, that is, in situations where the steering wheel occludes the driver's face. The proportion of *Road* ↔ *Unknown* transitions constituted 14% of all glance transitions, indicating that most of these were actually *Road-WindShield* transitions that could not be captured with the one-camera *Manual-DMS* system.

With the *L4* setup (Figure 4, right), slightly fewer of the derived transitions where between the *Road* and other AOIs (93.5%). Since the duration proportion to the *Road* was higher in the automated drives compared to the manual drives (see Figure 3, right), this suggests that glances to the *Road* were longer in automated drives, which is also confirmed in Figure 5 (right). In contrast to the manual drives, participants glanced more frequently between the *Road* and the *RightWindow* (43%), that is, the window next to them. *Road* ↔ *WindShield* glances accounted for 24%, while glances between the *Road* and the *Cluster* constituted only about 10% in total, which suggest that participants only rarely checked the speed of the vehicle (which they did not have to do in automated mode, and it could have been awkward given that they were sitting in the passenger seat).

3.3.3 | The effect of intoxication on glances

The effect of intoxication was assessed by performing LMER analyses with camera setup as a categorical fixed effect (with the *Manual* setup as reference level) and ground truth BrAC (see Section 3.1), nested within camera setup, as a continuous predictor. Random effects were modelled with random intercepts for participants and with by-participant random slopes for the effects of the setup and the intoxication level. In cases where the model failed to converge, the model was reduced to random intercept only.[11] Results derived with the reduced model are denoted with "†". Before running LMER, for each of the predictors, Tukey's fence method [39] was used to remove outliers within each camera setup and intoxication condition. The alpha-level was set to $0.05/18 = 0.00278$ to account for multiple models. Since the amount of data loss and *Unknown* glance events reflect camera setup rather than actual behaviour, *Unknown* glance events were excluded from the LMER analyses, and metrics were adjusted for the amount of data loss, where relevant.

Figure 5 shows the average glance rate (number of glances per minute with drive duration adjusted for data loss) and the average glance duration. Overall, in *Manual* setup, participants made fewer and longer glances as the intoxication level increased, where the glance rate decreased from an estimated 27 glances/min in sober drives to around 15 glances/min as the BrAC level increased to 1‰ ($p < 1e-4$, see Table 3 and blue line in the left of Figure 5). At the same time, glance durations increased from 2.43 to 4.6 s ($p < 1e-4$, Table 3 and blue line in the right of Figure 5). With increasing intoxication level the glance duration proportion to the *Road* increased an estimated 4.14% ($p = 1.43e-3$), meaning that in manual test track drives intoxicated drivers tended to glance between different AOIs less and instead looked longer at the road. This came at the expense

[11] The following formula was used: $lmer(measure \sim camera.setup + BrAC : camera.setup + (1 + camera.setup + BrAC | driverid)$, however, some models failed to converge, therefore model complexity for failing cases was reduced to random intercept only, that is, $lmer(measure \sim camera.setup + BrAC : camera.setup + (1|driverid)$.

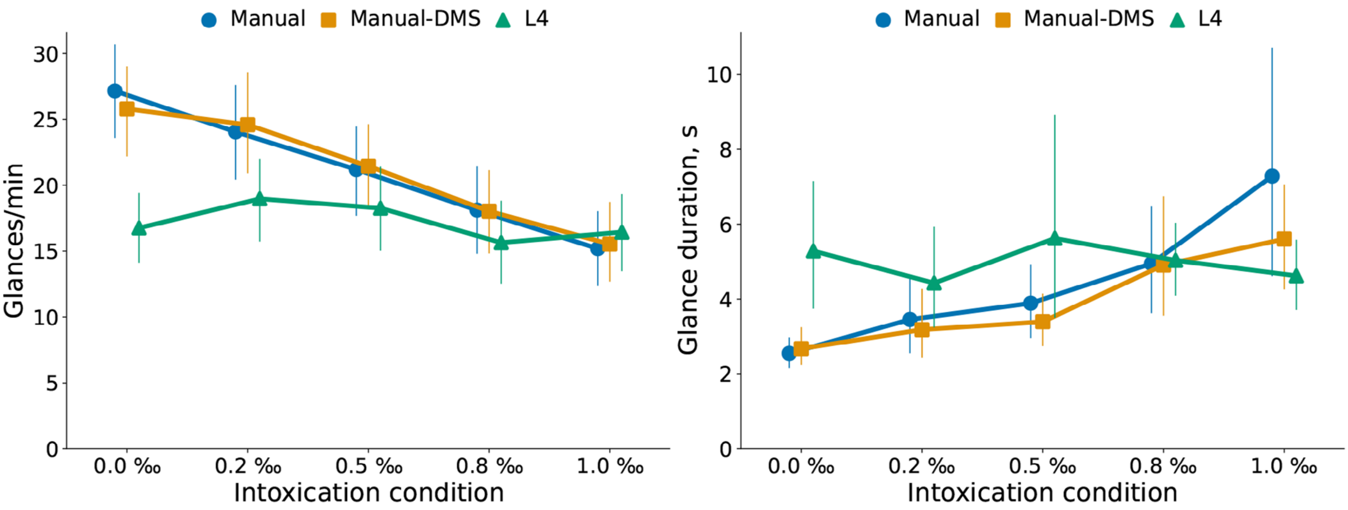

FIGURE 5 Left – average number of glances per minute (adjusted for the amount of data loss), right – average glance duration for different intoxication conditions with different camera setups. *Unknown* glance events were excluded. Error bars show 95% confidence interval.

TABLE 3 Linear mixed-effect regression model results for glance rate and glance duration. Significant results are in bold.

	Glance rate			**Glance duration**		
Predictors	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***
(Intercept)	26.79***	15.54	**9.26e-18**	2.43***	11.72	**1.45e-14**
Manual-DMS	−0.29	−0.35	7.24e-01	−0.15	−0.89	3.75e-01
L4	−9.10***	−6.11	**1.57e-07**	1.33**	3.96	**2.58e-04**
Manual:BrAC	−12.27***	−7.41	**1.15e-09**	2.15***	4.93	**1.17e-05**
Manual-DMS:BrAC	−11.00***	−6.51	**2.32e-08**	2.39***	5.43	**1.98e-06**
L4:BrAC	−1.47	−0.87	3.88e-01	1.18	2.70	9.61e-03
ICC			0.83			0.68
N drivers			35			35
Observations			491			457
Marginal R^2/conditional R^2			0.132 / 0.849			0.217 / 0.748

$^{*}p < 0.00278$; $^{**}p < 0.001$; $^{***}p < 1e-4$.

of 1.68% less glancing to the *WindShield* ($p < 0.001$) and 2.34% less glancing to the *Cluster* ($p < 1e-4$, Table 4 and blue lines in Figure 6) as well as a significantly lower transition rate between the *Road* and the *Cluster* (−5.76, $p < 1e-4$) and the *Road* and the *WindShield* (−2.81, $p < 1e-4$, Table 5 and top of Figure 7).

The same general trends were found for the *Manual-DMS* as for the *Manual* setup. After adjusting for data loss, the glance rate was slightly, but not significantly, lower compared to that of the *Manual* setup (−0.29 glances, $p = 7.24e-1$, see also yellow line in Figure 5). The average glance duration was also the same. The effect of BrAC was somewhat different compared to the *Manual* setup, nonetheless highly significant for both glance rate and glance duration (Table 3). Glance duration proportion and transition measures were however severely affected by high data loss in the *Manual-DMS* setup. First, a higher glance proportion was found to the *Road* (4.7%, $p < 1e-4$), and a corresponding lower proportion was found to the *WindShield* (−4.9%, $p < 1e-4$), as well as significantly fewer *Road-WindShield* transitions in sober condition. In addition, the significant effect of BrAC level on glance proportion to both the *Road* and the *WindShield* AOIs could not longer be measured, although a reduction of transition rate between the two AOIs was still significant ($p < 0.001$, Table 5). In summary, estimates differed between the *Manual* and *Manual-DMS* setup in all metrics except glance duration proportion to the *Cluster* and transitions between the *Road* and the *Cluster*.

Contrary to the *Manual* setup, the effect of intoxication was not that obvious in the *L4* setup. When sober in automated mode, participants made an estimated 9 glances per minute less ($p < 1e-4$), which were an estimated 1.3 seconds longer ($p < 0.001$) compared to when they were driving manually. Drivers however looked 5% more to the road ($p < 0.001$) and only very little to the *Cluster* (≈ 1.2%, $p < 1e-4$) since they did not need to monitor speed of the vehicle (see Tables 3 and 4 and green

TABLE 4 Linear Mixed-Effect Regression model results for glance duration proportion to the *Road*, *Cluster*, and *WindShield* AOIs. Significant results are in bold. "†" denotes results derived with the random intercept only model [11].

	Road			Cluster†			WindShield†		
Predictors	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***
(Intercept)	83.43***	71.09	**5.49e-42**	7.01***	13.29	**3.76e-21**	6.66***	19.54	**1.96e-39**
Manual-DMS	4.70***	5.88	**7.01e-08**	0.35	0.75	4.51e-01	−4.91***	−12.81	**3.31e-32**
L4	4.96**	3.72	**5.54e-04**	−5.86***	−12.50	**7.50e-31**	−4.08***	−11.02	**3.96e-25**
Manual:BrAC	4.14*	3.36	**1.43e-03**	−2.34***	−4.21	**3.09e-05**	−1.68**	−3.90	**1.10e-04**
Manual-DMS:BrAC	2.84	2.24	2.88e-02	−2.27***	−3.95	**9.28e-05**	−0.70	−1.49	1.36e-01
L4:BrAC	−1.47	−1.15	2.54e-01	−0.24	−0.41	6.86e-01	−0.33	−0.72	4.73e-01
ICC			0.75			0.48			0.30
N drivers			35			35			35
Observations			478			477			483
Marginal R^2/conditional R^2			0.081 / 0.772			0.325 / 0.652			0.397 / 0.578

*$p < 0.00278$; **$p < 0.001$; ***$p < 1e-4$.

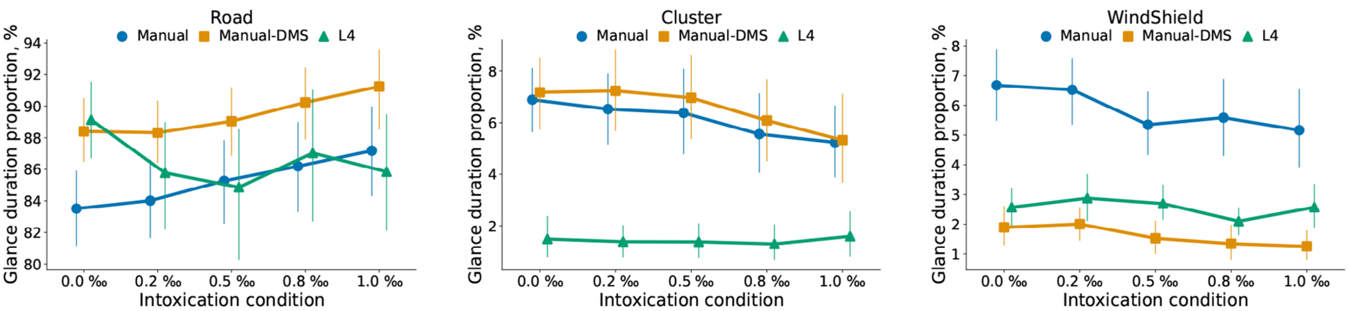

FIGURE 6 Average normalised glance duration proportion to the *Road*, *Cluster*, and *WindShield* AOIs for different intoxication conditions with different camera setups. Error bars show 95% confidence interval.

lines in Figures 5 and 6). None of the above mentioned parameters changed significantly when driving under the influence of alcohol, although we cannot reject the possibility that the glance proportion measures in the automated drives were affected by data loss just like the *Manual-DMS* setup, since data loss in the *L4* was, on average, more that six times that in *Manual* setup (Table 2).

The proportion of glance duration to the *WindShield* in the automated drives was significantly lower compared to the *Manual* setup (2.6% vs 6.66%, $p < 1e-4$) and did not change when intoxicated. The glance transition rate between the *Road* and the *WindShield* was however significantly reduced under the influence of alcohol ($p < 0.001$, Table 5). Note that this result is a direct consequence of the camera setup. In the automated drives, participants were sitting in the passenger seat on the right-hand side of the car, so when looking at traffic signs, participants would shift their gaze to the right, which in manual driving aligned with the *WindShield* AOI, while during automated driving, it was the *RightWindow*. Moreover, our test track had multiple sharp right turns and somewhat less sharp left turns (see Figure A1 in Appendix A), which also likely resulted in more extreme right gaze angles. Therefore, in automated drives, participants had little to no reason to look through the *WindShield* on their left and instead right-side glances aligned with the *RightWindow* AOI. This can be seen in Figure 8 (middle), where the share of glancing to the *RightWindow* was significantly higher compared to that in the *Manual* setup (an estimated 5.64% vs 0.79%, $p < 1e-4$, Table 6), as well as in the glance transition rate from the *Road* to the *RightWindow* – an estimated 7.98 glances/min vs 1.18 glances/min ($p < 1e-4$, Table 5 and Figure 7). Also, similarly to the glances to the *Road* and the *WindShield* in the *Manual* setup, when driving intoxicated in automated mode, glance duration proportion to the *RightWindow* decreased, although insignificantly, yet the decrease in *Road* ↔ *RightWindow* glance transition rate was highly significant (-3.77 glances/min, $p < 1e-4$, see Table 5), further indicating that what was *Road* ↔ *WindShield* glances in manual drives, became *Road* ↔ *RightWindow* glances when driving in automated mode.

In automated driving, glance duration proportion to the *LeftWindow* increased by 2.59% ($p < 1e-4$) with a BrAC level of 1‰, that is, as drivers became more intoxicated, they looked more to the direction of the test leader sitting next to them on the left. Very similar, yet opposite and not so clearly expressed, glance behaviour was observed in the manual drives. The glance

TABLE 5 Linear mixed-effect regression model results for omnidirectional glance transition rate between the *Road* and *Cluster*, *WindShield*, *LeftWindow* and the *RightWindow* AOIs. Significant results are in bold. "†" denotes results derived with the random intercept only model [11].

	Road-Cluster†			Road-WindShield			Road-LeftWindow			Road-RightWindow†		
Predictors	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***
(Intercept)	12.98***	13.89	**2.12e-21**	5.97***	15.55	**1.32e-18**	4.61***	12.79	**8.14e-16**	1.18*	3.25	**1.41e-03**
Manual-DMS	−0.90	−1.17	2.44e-01	−3.35***	−10.94	**2.81e-18**	−1.17***	−4.56	**9.78e-06**	−0.26	−0.62	5.36e-01
L4	−11.55***	−14.48	**1.86e-39**	−2.16***	−5.30	**2.04e-06**	−3.44***	−9.42	**1.54e-13**	6.69***	16.04	**1.88e-46**
Manual:BrAC	−5.76***	−6.29	**7.39e-10**	−2.81***	−7.24	**3.05e-10**	−3.57***	−9.31	**4.30e-14**	0.59	1.18	2.37e-01
Manual-DMS:BrAC	−6.05***	−6.59	**1.23e-10**	−1.45**	−3.69	**4.05e-04**	−2.84***	−7.34	**1.98e-10**	0.14	0.27	7.84e-01
L4:BrAC	0.11	0.11	9.14e-01	−1.53**	−3.94	**1.77e-04**	1.36**	3.47	**8.36e-04**	−3.77***	−7.66	**1.04e-13**
ICC			0.52			0.66			0.65			0.23
N drivers			35			35			35			35
Observations			504			508			503			508
Marginal R^2/conditional R^2			0.315 / 0.672			0.262 / 0.745			0.200 / 0.723			0.463 / 0.586

*$p < 0.00278$; **$p < 0.001$; ***$p < 1e-4$.

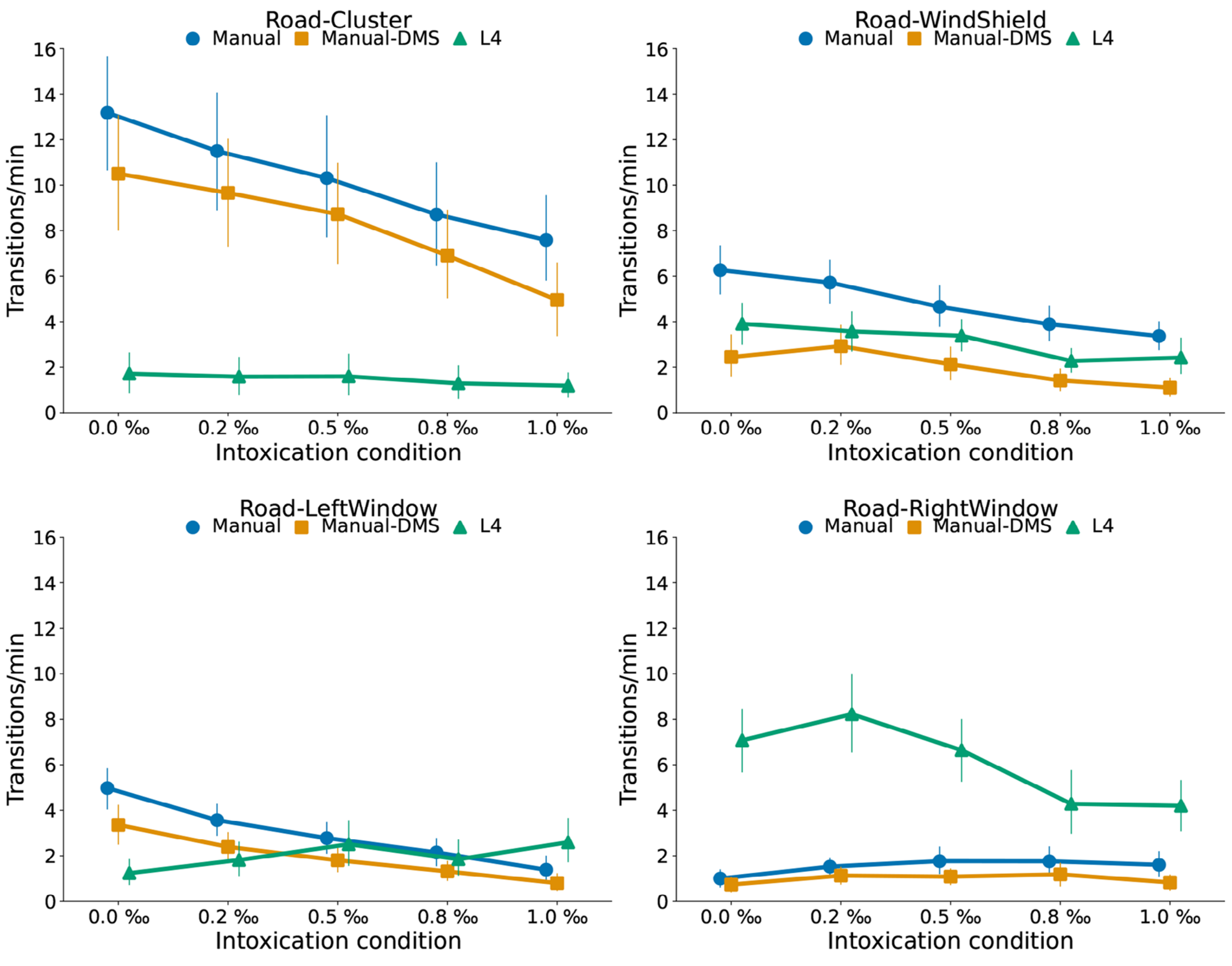

FIGURE 7 Glance transition rate between the *Road* and selected AOIs for different intoxication levels with different camera setups. Error bars show 95% confidence interval.

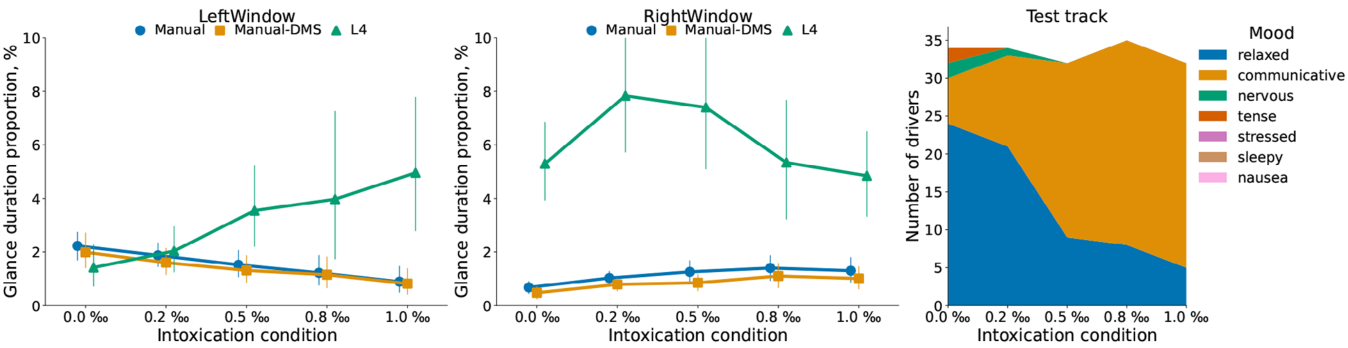

FIGURE 8 Average normalised glance proportion to the *LeftWindow* and the *RightWindow* AOIs for different intoxication conditions with different camera setups. Error bars show 95% confidence interval.

TABLE 6 Linear Mixed-Effect Regression model results for glance duration proportion to the *LeftWindow* and the *RightWindow* AOIs. Significant results are in bold.

	LeftWindow			RightWindow		
Predictors	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***
(Intercept)	2.07***	9.83	**3.05e-13**	0.79**	3.80	**2.38e-04**
Manual-DMS	−0.46	−2.25	**2.49e-02**	−0.31	−1.05	2.96e-01
L4	−0.79	−1.87	6.78e-02	4.85***	10.25	**4.96e-14**
Manual:BrAC	−1.50***	−4.81	**7.26e-06**	0.70	1.67	9.86e-02
Manual-DMS:BrAC	−1.02*	−3.16	**2.19e-03**	0.73	1.64	1.04e-01
L4:BrAC	2.59***	7.82	**6.80e-12**	−1.27	−2.82	**5.84e-03**
ICC			0.63			0.44
N drivers			35			35
Observations			467			474
Marginal R^2/conditional R^2			0.195 / 0.702			0.461 / 0.698

*$p < 0.00278$; **$p < 0.001$; ***$p < 1e-4$.

duration proportion to the *LeftWindow* AOI (i.e. window next to the driver) dropped (-1.5%, $p < 1e-4$), while proportion to the direction of the test leader sitting in the passenger seat (i.e. the *RightWindow* AOI) increased approximately twofold (albeit this effect was not found to be significant ($p = 9.86e-02$), see Table 6). The same patters were observed for glance transitions (bottom of Figure 7 and Table 5). This finding can be attributed to engagement in social interaction. After each of the drives, the test leader annotated the participant's mood. Figure 8 (right) shows the distribution of mood annotations across all intoxication conditions. In the sober condition, the vast majority of drivers were deemed to be relaxed. However, as the intoxication level increased, more and more participants were reported as "communicative" (see blue and yellow areas in the right of Figure 8).

3.4 | Psychophysiological measures

Results from the glance analyses showed that glances were very much dependent on the driving mode, the camera setup, and most likely also with learning effects and familiarity with the task. This makes it hard to pinpoint exactly which effects are due to intoxication. We therefore continued by exploring lower-level psychophysiological measures of fixations, saccades, and eye blinks. To evaluate general changes in drivers' state, we use all fixations and blinks independent of where the drivers are looking; however, to disentangle behaviours influenced by car layout and other environmental factors (such as talking to the test leader, looking at the mirrors and similar), we separately analyse only fixations and saccades within the *Road* AOI, thus evaluating only road scanning behavior.

3.4.1 | Intoxication effects on fixations and saccades

Fixation rate and average fixation duration results (Figure 9) show that as the intoxication level increased, in all camera setups, drivers were found to make fewer but longer fixations (all interactions were highly significant with $p < 1e-4$). For example, in the *Manual* setup, the fixation rate dropped from

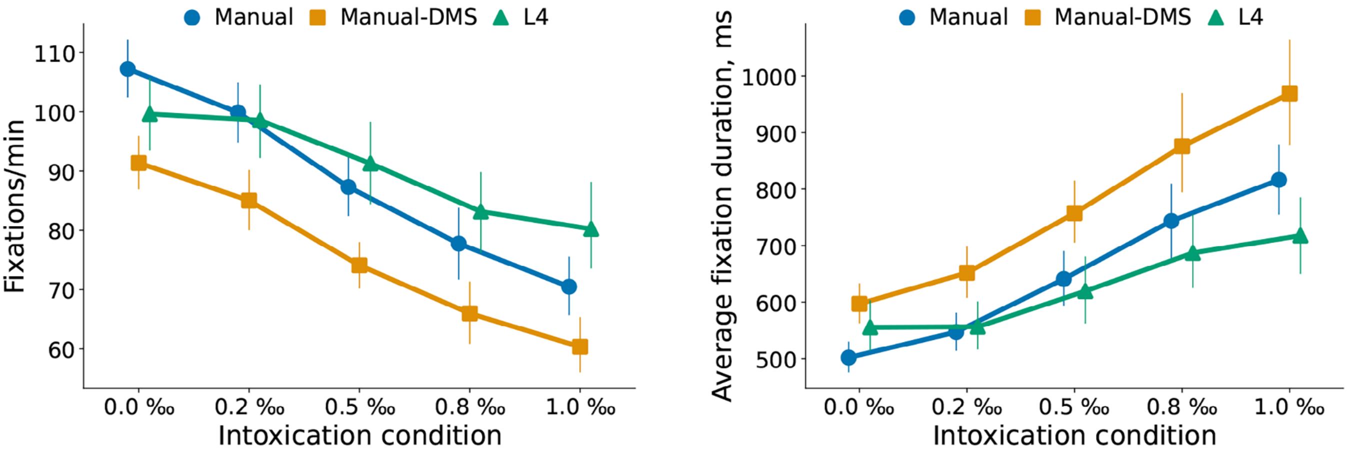

FIGURE 9 Average fixation rate (adjusted for data loss) and average fixation duration for different intoxication conditions with different camera setups. Error bars indicate 95% confidence interval.

TABLE 7 Linear Mixed-Effect Regression model results for fixation rate and duration, mean saccade amplitude and fixation $\sqrt{BCEA}$. Significant results are in bold. "†" denotes results derived with the random intercept only model [11].

	Fixation rate†			Fixation duration			Saccade amplitude (road)			Fixation $\sqrt{BCEA}$ (Road)		
Predictors	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***
(Intercept)	106.99***	44.21	**2.29e-48**	486.28***	35.78	**9.58e-36**	5.84***	47.05	**9.46e-40**	13.60***	33.47	**4.39e-32**
Manual-DMS	−15.74***	−7.97	**1.40e-14**	93.05***	6.95	**6.59e-11**	0.93***	7.68	**7.34e-12**	−0.09	−0.24	8.12e-01
L4	−5.41	−2.83	4.93e-03	41.22	2.30	2.46e-02	0.92***	5.29	**2.24e-06**	1.24	2.17	3.52e-02
Manual:BrAC	−38.29***	−17.12	**8.26e-51**	334.22***	10.48	**1.81e-13**	−0.87***	−5.38	**7.09e-07**	−0.34	-0.78	4.36e-01
Manual-DMS:BrAC	−32.40***	−13.38	**1.59e-34**	364.45***	11.22	**7.83e-15**	−0.81***	−4.80	**6.29e-06**	0.09	0.21	8.37e-01
L4:BrAC	−22.50***	−9.56	**8.26e-20**	194.23***	5.99	**2.94e-07**	−0.29	−1.74	8.51e-02	0.69	1.56	1.21e-01
ICC			0.56			0.61			0.60			0.60
N drivers			35			35			35			35
Observations			483			477			486			490
Marginal R^2/conditional R^2			0.408 / 0.741			0.577 / 0.834			0.298 / 0.721			0.086 / 0.632

*$p < 0.00278$; **$p < 0.001$; ***$p < 1e-4$.

an estimated 107 fixations per minute in sober drives down to an estimated 69 fixations per second when at 1.0‰ BrAC. The average fixation duration increased from 486 ms to 821 ms (see blue line in Figure 9 and Table 7). These results indicate that the visual system needs more time to process visual information when under the influence of alcohol. With the *Manual-DMS* setup, the derived fixation metrics were substantially different (see yellow line in the top of Figure 9) compared to when extracted from the *Manual* setup, with significantly fewer but longer fixations ($p < 1e-4$). The *Manual-DMS* recordings were considerably noisier (2.0° vs 1.18° RMS, Table 2), leading to longer fixations since smaller saccades could not be detected. As a consequence, the mean saccade amplitude became larger. As an example, the derived saccade amplitude inside the *Road* AOI (Figure 10, left) was 0.93 degree larger ($p < 1e-4$) with the *Manual-DMS* setup as compared to the *Manual* setup (Table 7). The effect of intoxication could nevertheless still be measured with the *Manual-DMS* setup ($p < 1e-4$ for all above mentioned metrics).

During automated driving, neither fixation rate nor average fixation duration differed significantly compared to the *Manual* setup, while the mean saccade amplitude inside the *Road* AOI was significantly higher ($p < 1e-4$) and similar to that found in the *Manual-DMS*. Fixations, but not saccade amplitudes, were significantly affected by alcohol intoxication, with longer fixations ($p < 1e-4$). Since the noise level in the *Manual* and *L4* setups were substantially different (1.18° vs 1.83° RMS, see Table 2), it cannot be concluded that the fixations were actually the same in the two setups.

The square root of the Bivariate Contour Ellipse Area ($\sqrt{BCEA}$) [46, 47] of fixation positions inside the *Road* AOI was used as a proxy of attention span (right of Figure 10). An estimated BCEA in the *Manual* setup was 13.6 degrees, which was around 1/3 of the horizontal span of the *Road* AOI. $\sqrt{BCEA}$ was 1.24 degrees larger (not significant, $p = 3.52e-02$) in the *L4* compared to in the *Manual* setup.

The most surprising result was that no significant intoxication effects could be found on attention span in any of the camera

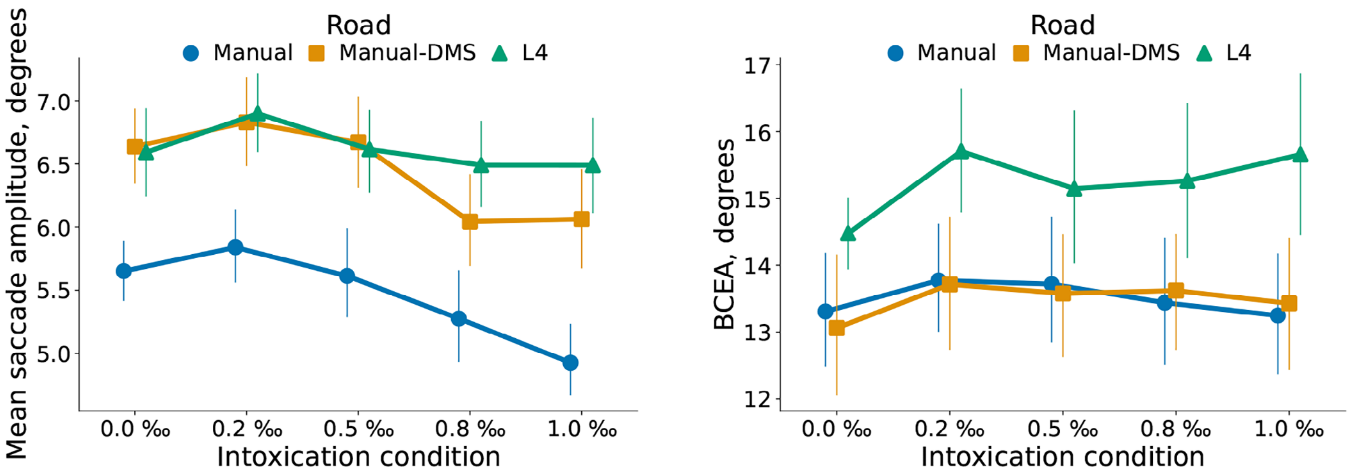

FIGURE 10 Mean saccade amplitude and fixation distribution within the *Road* AOI for different intoxication conditions with different camera setups. Error bars indicate 95% confidence interval.

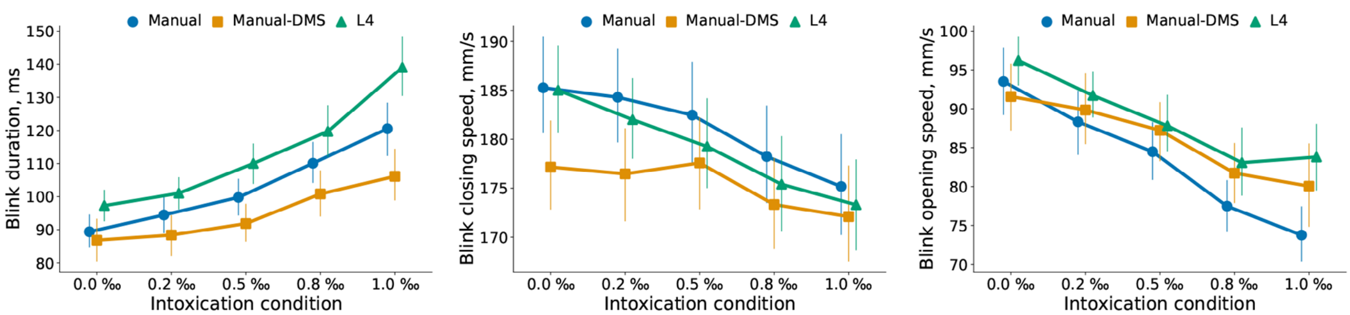

FIGURE 11 Blink dynamics measures for different intoxication conditions with different camera setups. Error bars show 95% confidence interval.

setups, indicating that even when highly intoxicated, the participants scanned the forward road in a similar way as in the sober condition. At the same time, in the *Manual* setup from intoxicated drivers, saccade amplitudes became 0.87 degrees smaller ($p < 1e-4$) while they did not change significantly in the *L4*.

3.4.2 | Intoxication effect on blink dynamics

Figure 11 shows the average blink duration and blink closing/opening velocities with different camera setups and intoxication conditions. As expected, blink durations became longer (24–39 ms with different camera setups, $p < 1e-4$), the blink closing speed decreased (by 6.63–13 mm/s, $p < 1e-3$), and the blink opening speed was reduced (by 13.7–19.8 mm/s, $p < 1e-4$), see Table 8. Somewhat longer blinks with somewhat slower closing and faster opening speeds were found in the *L4* setup, but none of these differences were significant. Except for the blink closing speed, which was an estimated 7.5 mm/s slower ($p < 1e-4$) in the *Manual-DMS*, there were no significant differences in the blink data from the *Manual* and *Manual-DMS* setups.

4 | DISCUSSION

This paper investigated how various behavioural and psychophysiological eye tracking measures were affected by driver intoxication, and whether these effects can be measured using a single-camera DMS, as well as when the driver is out-of-the-loop during L4 automated driving. 35 participants performed controlled manual and automated L4 test track drives sober and then at four increasing intoxication levels reaching a BrAC around 1‰. Three separate eye tracking camera setups were used to capture participants' gaze and blink data: a five-camera system on the dashboard and in the interior of the vehicle, a one-camera DMS mounted on the steering column, and a one-camera system mounted on the dashboard at the passenger side (to record the driver during the automated driving sessions).

4.1 | Effects of intoxication on behavioural and psychophysiological eye tracking measures

The glance duration and transition analyses showed that in manual drives with increasing intoxication level, the average glance rate decreased, while glance duration increased approximately

TABLE 8 Linear mixed-effect regression model results for blink dynamics measures. Significant results are in bold.

	Blink duration			Blink closing speed			Blink opening speed		
Predictors	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***	**Estimates**	**Statistic**	***p***
(Intercept)	87.47***	33.31	**1.27e-30**	186.24***	74.21	**6.70e-41**	92.96***	50.77	**1.49e-35**
Manual-DMS	−3.60	−1.82	7.14e-02	−7.46***	−5.86	**8.83e-08**	0.03	0.02	9.81e-01
L4	6.76	3.01	3.56e-03	−1.31	−1.10	2.74e-01	2.23	1.71	9.18e-02
Manual:BrAC	31.44***	9.44	**3.73e-13**	−10.68***	−6.37	**2.04e-08**	−19.83***	−12.72	**9.02e-19**
Manual-DMS:BrAC	23.96***	7.00	**2.25e-09**	−6.63**	−3.82	**2.77e-04**	−14.23***	−8.88	**6.39e-13**
L4:BrAC	39.18***	11.57	**7.78e-17**	−13.01***	−7.59	**8.91e-11**	−13.71***	−8.62	**2.25e-12**
ICC			0.67			0.86			0.83
N drivers			35			35			35
Observations			491			490			493
Marginal R^2/Conditional R^2			0.415 / 0.809			0.088 / 0.874			0.240 / 0.869

*$p < 0.00278$; **$p < 0.001$; ***$p < 1e-4$.

twofold. This was driven by two factors. First, participants made fewer and longer fixations, thus increasing the duration of glances, which indicates that intoxication negatively affected the visual system of the drivers who therefore need more time to process visual information [48]. Second, participants tended to increase their glance share to the forward road, mostly at the expense of glances to the sides and to the instrument cluster, leading to reduced situational awareness. With intoxication, drivers glanced between different AOIs less and instead glanced longer at the road. Similar results have been reported before [19, 49].

In the automated drives, compared to the manual drives, participants made fewer but longer glances. However, the glance metrics did not vary with the intoxication level in the automated drives. This indicates that higher-level measures such as glances are too affected by the driving environment/situation to be useful for intoxication detection. In contrast, participants made fewer but longer fixations when intoxicated, regardless of driving mode, showing that lower-level processes have better potential to be used for estimating a driver's state in different settings.

Intoxication also affected blink dynamics across driving modes, with longer and slower blinks for increasing intoxication levels. An issue if blink metrics are to be used for intoxication detection is that blink dynamics also vary with driver sleepiness [see e.g. [50–52]]. In this study, drivers were asked to evaluate their sleepiness level using the *Karolinska Sleepiness Scale* [30] and were found to become sleepier throughout the trials, but on a low level (alert or rather alert on average, see Figure D1 in Appendix D). We therefore conclude that in our study changes found in blink dynamics were largely due to intoxication and not due to increased sleepiness. In a DMS application for automatic intoxication detection, it will however be difficult to disentangle whether increased blink durations are due to sleepiness or intoxication.

Despite intoxication causing a narrowed glance distribution with more glances to the *Road* AOI, which is in line with [49], we did not find any significant intoxication effects on attention span (evaluated as dispersion of fixations inside the *Road* AOI) in any of the driving modes. This indicates that even though situational awareness may be worsened due to reduced scanning of more peripheral AOIs, the scanning of the actual road remains intact, which may be a contributing reason to why non-complex behaviours such as lane keeping are rather unaffected by alcohol intoxication [53]

Finally, the results show that with increasing intoxication level, drivers become more likely to engage in non-driving related tasks [27 see also]. Here, intoxication resulted in more talking with the test leader sitting next to the participants. In the manual drives, the glance rate to more peripheral areas (such as the mirrors, road signs, upcoming curves, or the scenery) decreased, while glances towards the test leader sitting in the passenger seat increased, albeit not significantly (see Table 5).

4.2 | Data quality considerations

Available data analysis methods, metrics and therefore result interpretations as well as possible conclusions that can be drawn from an eye tracking study depend greatly on data quality [54]. In this study, three different setups were used to record eye tracking data, why data quality considerations are especially important.

For example, glance duration proportion measures, especially the normalized glance proportion, is very sensitive to data loss. In the *Manual-DMS* setup using a camera mounted on the steering column, the camera got occluded in specific driving scenarios such as turning. As a consequence, the glance proportion to the *WindShield* AOI (the area where drivers are likely to look when turning right) appeared to be three times lower compared to that the five-camera *Manual* setup, despite being recorded at the same time. There were also significant differences in the proportion of glances to the *Road* as well as in *Road* ↔ *WindShield* and *Road* ↔ *LeftWindow* transitions. Also, some metrics lost their relation with intoxication when measured with the *Manual-DMS* setup. Similarly, it is not possible to say if the differences and similarities found in glance metrics derived from the *Manual* as compared to the *L4* setup were due to driving mode or due to different data losses/quality in the different camera setups. One thing that can be said is that the data loss was on average more than six times higher in the *L4* setup compared to the *Manual* setup.

The impact of different noise levels in the three camera setups was mostly noticeable in the fixation and saccade measures, but glance based metrics were also affected since glances are derived from fixations. There were about 15% fewer fixations, which were about 20% longer, when the fixations were calculated from data in the *Manual-DMS* setup as compared to the *Manual* setup. At the same time, the analysed saccade amplitudes were about 1 degree larger. This is a well known effect of noise in gaze data. As the noise level increases, smaller saccades cannot be detected, leading to separate fixations being merged into longer fixations. This effectively increases the average fixation duration, as well as the mean saccade amplitude [55 fig. 4, [32], fig. 11 and 12]. The general effect of intoxication nevertheless was similar for all mentioned metrics and could be measured in all three camera setups.

The noise level in the gaze signal was more than 50% higher in the *L4* setup compared to in the five-camera *Manual* setup. Still, there were no significant differences in the fixation metrics between the setups, where only saccade amplitudes differed. It is very likely that fixations were actually shorter during automated driving, but because of the higher noise levels in the *L4* gaze data, the fixations were merged into longer fixations which then happened to align with those in the *Manual* setup. Further, comparing saccade amplitudes derived from the *L4* and *Manual-DMS* setups, we speculate that the amplitudes were in fact very similar in manual and automated driving since the noise levels in the two were similar, so the derived saccade amplitudes should also be similar. However, in the automated drives, no intoxication effect was found in the saccade amplitudes, why more studies are needed where manual and automated driving modes are compared using identical experimental setups.

Finally, data loss and noise analysis (Figure C1 in Appendix C) showed that with increasing intoxication level, driver behaviour changed to such an extent that it generally increased the noise levels of both single-camera systems. The behavioural changes, such as more interaction with the test leader and thus more extreme glances, and their associated increase in noise levels made it difficult to capture some of the investigated eye tracking metrics with a one-camera system, something which warrants careful interpretation of the results.

4.3 | Limitations

There are several limitations that must be considered when interpreting the results from this study. First, the automated condition was emulated by a test leader driving the car while the participant sat in the passenger seat. Ideally, the automated driving sessions should of course be driven in an actual automated vehicle, or in a more sophisticated Wizard-of-Oz car where the participant could remain seated on the left side [56]. As it was, sitting in the passenger seat affected visual scanning, which is not ideal in an eye tracking study. Also, the manual drives always preceded the automated drives and were not counterbalanced. Should we repeat the experiment we would ensure that participants got a practice run also in automated mode, and that the order (manual versus automated) would be counterbalanced. The choice to always start with manual driving was based on a prioritization of manual driving data. In case participants started to feel unwell due to intoxication or motion sickness in the automated drives, then at least data from the manual drive would be available.

Second, because of the limited space on the dashboard of the car, it was not possible to use a multi-camera eye tracking setup on the right side as well. Instead, only a one-camera system was used in the automated driving condition. In addition, there were slight differences between the one-camera system used in the *Manual-DMS* setup and the one-camera system used in the *L4* setup. Consequently, it was not possible to make a one-to-one comparison between the manual and automated driving conditions.

Third, no placebo group was used to control for familiarity, making it difficult to untangle learning effects from intoxication effects. This was a due to budget constraints in the project, where a larger number of participants in the alcohol group was prioritised over a placebo/control group. A follow-up study with a placebo control group is recommended.

Finally, the intoxication conditions were not counterbalanced and the level of intoxication was increased incrementally. Another approach would have been to target the highest level of 1‰ BrAC directly and then run the tests as the BrAC level wore off over time. An advantage with the latter approach is that it would have been easier to pinpoint the different target levels. However, the former approach was chosen since it maximizes the possible number of drives from each participant—if a participant does not feel well due to a high BrAC, we have already collected the drives on the lower levels. One may also assume that participants are generally more motivated and positive while on the incremental side of the curve, while they might feel less motivated in the process of sobering up. With a rate of a BrAC reduction of around 0.1‰ per hour, we would have needed to keep the participants at the test site for a disproportionally long time. In general, all these limitations makes it difficult to disentangle effects from intoxication, driving mode, camera setup, and learning.

5 | CONCLUSIONS

Alcohol intoxication affects how drivers look at the road and surrounding areas, as well as lower-level features, such as fixations, saccades, and blinks, in both manual and automated L4 driving modes. We found that drivers focused more on the forward road when intoxicated, at the expense of reduced scanning of peripheral areas. While higher-level glance measures gave insights in how drivers change their behaviour when intoxicated, we further found that such metrics vary considerably with driving context and driving/automation mode. Lower-level psychophysiological measures—fixation rate and duration, blink duration, and blink opening/closing speed—seem to reflect the driver's state better, and the effect of intoxication was found to be more independent of driving mode. However, depending on hardware configuration and therefore noise in eye tracking signals, absolute values of these measures differ from system to system; why one must be cautious when using them for alcohol intoxication detection system development.

Not all effects of alcohol intoxication could be measured when using certain hardware configurations, making automatic intoxication detection a very challenging task. The results demonstrate the importance of taking data quality into account when interpreting the results, as some metrics might be affected more than others in ways that are not always predictable. Consequently, this means that data from different studies with different sensor setups might not be comparable without additional context. In real DMS applications, this means that lab findings based on higher quality eye tracking systems might not be directly applicable to production settings because of hardware limitations. Even if general trends of the metrics can still be captured when using different camera types and placement, the absolute value estimates might be off, thus requiring accounting for data loss or noise. Some measures, especially the proportional metrics, should therefore be interpreted with caution. Similarly, one should be very careful when comparing, for example, fixation and saccade metrics that come from data with different noise levels.

Given findings and limitations of the current study, future work should focus on investigating driving in more realistic driving conditions, for example, simulate other traffic, pedestrians, and every day driving situations, such as stopping or driving through intersections. We also plan to include a placebo group to control for route familiarity and learning effects, and use identical eye tracking sensor setups in different driving conditions to enable side-by-side comparisons.

AUTHOR CONTRIBUTIONS

Raimondas Zemblys: Conceptualization; data curation; formal analysis; methodology; software; visualization; writing—original draft; review & editing. **Christer Ahlström**: Conceptualization; investigation; methodology; writing—review & editing. **Katja Kircher**: Funding acquisition; investigation; methodology; project administration; supervision. **Svitlana Finér**: Funding acquisition; project administration; resources; writing—review & editing.

ACKNOWLEDGEMENTS

We would like to thank the VTI data collection team and the Smart Eye Data Acquisition, System Development, Research and Advanced Engineering, and Operations teams for their help and support with data collection. This work was partly funded by the Swedish Strategic Vehicle Research and Innovation Programme (VINNOVA FFI; Grant No. 2019-05834, https://www.vinnova.se/p/fit2drive).

CONFLICT OF INTEREST STATEMENT

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Raimondas Zemblys and Svitlana Finér are employed by the eye tracking company Smart Eye AB, Gothenburg, Sweden. All authors declare that there are no additional conflicts of interest.

DATA AVAILABILITY STATEMENT

The authors do not have permission to share data. The experiment was not preregistered.

CONSENT FOR PUBLICATION

All the authors mentioned in the manuscript have agreed for authorship, read and approved the manuscript, and gave consent for submission and subsequent publication of the manuscript.

DECLARATION OF GENERATIVE AI IN SCIENTIFIC WRITING

Authors did not use AI and AI-assisted technologies in the writing process.

ORCID

Raimondas Zemblys https://orcid.org/0000-0002-5316-7491
Christer Ahlström https://orcid.org/0000-0003-4134-0303
Katja Kircher https://orcid.org/0000-0002-1849-9722
Svitlana Finér https://orcid.org/0000-0002-1931-2448

REFERENCES

1. Modijefsky, M., Janse, R., Spit, W., Jankowska-Karpa, D., Buttler, I., Eikefjord, B.: Prevention of driving under the influence of alcohol and drugs. Final report, Publications Office of the European Union, Luxembourg (2022)
2. Stewart, T.: 2022. Overview of motor vehicle crashes in 2020. National Highway Traffic Safety Administration. DOT HS 813 266. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/813266
3. Vissers, L., Houwing, S., Wegman, F.: Alcohol-Related Road Casualties in Official Crash Statistics. ITF: International Transport Forum, Paris (2018)
4. Voas, R.B., Torres, P., Romano, E., Lacey, J.H.: Alcohol-related risk of driver fatalities: An update using 2007 data. J. Stud. Alcohol Drugs 73(3), 341–350 (2012)
5. Ranney, T.A., Gawron, V.J.: The effects of pavement edgelines on performance in a driving simulator under sober and alcohol-dosed conditions. Human Fact. 28(5), 511–525 (1986)
6. Huemer, A.K., Vollrath, M.: Alcohol-related impairment in the Lane Change Task. Accid. Anal. Prev. 42(6), 1983–1988 (2010)
7. Helland, A., Jenssen, G.D., Lervåg, L.E., Westin, A.A., Moen, T., Sakshaug, K., et al.: Comparison of driving simulator performance with real driving after alcohol intake: A randomised, single blind, placebo-controlled, cross-over trial. Accid. Anal. Prev. 53, 9–16 (2013)
8. Fillmore, M.T., Blackburn, J.S., Harrison, E.L.: Acute disinhibiting effects of alcohol as a factor in risky driving behavior. Drug Alcohol Depend. 95(1-2), 97–106 (2008)
9. Mitchell, M.C.: Alcohol-induced impairment of central nervous system function: behavioral skills involved in driving. J. Stud. Alcohol Suppl. 46(10), 109–116 (1985)
10. Fillmore, M.T.: Acute alcohol-induced impairment of cognitive functions: Past and present findings. Int. J. Disability Human Develop. 6(2), 115–126 (2007)
11. Garrisson, H., Scholey, A., Ogden, E., Benson, S.: The effects of alcohol intoxication on cognitive functions critical for driving: A systematic review. Accid. Anal. Prev. 154, 106052 (2021)

12. Euro, NCAP: Euro NCAP Vision 2030: A Safer Future for Mobility. https://www.euroncap.com/en/press-media/press-releases/euro-ncap-vision-2030-a-safer-future-for-mobility/
13. Ferguson, S.A., Zaouk, A., Strohl, C.: Driver alcohol detection system for safety (DADSS): Background and rationale for technology approaches. SAE Technical Paper 2010-01-1580 (2010)
14. Ljungblad, J., Hök, B., Allalou, A., Pettersson, H.: Passive in-vehicle driver breath alcohol detection using advanced sensor signal acquisition and fusion. Traffic Inj. Prev. 18(sup1), S31–S36 (2017)
15. Webster, G.D., Gabler, H.C.: Feasibility of transdermal ethanol sensing for the detection of intoxicated drivers. In: Annual Proceedings/Association for the Advancement of Automotive Medicine, vol. 51, p. 449. Association for the Advancement of Automotive Medicine, Des Plaines, IL (2007)
16. Rahman, H., Barua, S., Ahmed, M.U., Begum, S., Hök, B.: A case-based classification for drivers' alcohol detection using physiological signals. In: Internet of Things Technologies for HealthCare: Third International Conference, HealthyIoT 2016, Revised Selected Papers, vol. 3, pp. 22–29. Springer, Cham (2016)
17. Koch, K., Maritsch, M., Van Weenen, E., Feuerriegel, S., Pfäffli, M., Fleisch, E., et al.: Leveraging driver vehicle and environment interaction: Machine learning using driver monitoring cameras to detect drunk driving. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–32. ACM, New York (2023)
18. Shirazi, M.M., Rad, A.B.: Detection of intoxicated drivers using online system identification of steering behavior. IEEE Trans. Intell. Transp. Syst. 15(4), 1738–1747 (2014)
19. Maurage, P., Masson, N., Bollen, Z., D'Hondt, F.: Eye tracking correlates of acute alcohol consumption: A systematic and critical review. Neurosci. Biobehav. Rev. 108, 400–422 (2020)
20. Nawrot, M., Nordenstrom, B., Olson, A.: Disruption of eye movements by ethanol intoxication affects perception of depth from motion parallax. Psycholog. Sci. 15(12), 858–865 (2004)
21. Roche, D.J., King, A.C.: Alcohol impairment of saccadic and smooth pursuit eye movements: impact of risk factors for alcohol dependence. Psychopharmacology 212, 33–44 (2010)
22. Schmitt, K.U., Lanz, C., Muser, M.H., Walz, F., Schwarz, U.: Saccadic eye movements after low-dose oral alcohol exposure. J. Forensic Legal Med. 20(7), 870–874 (2013)
23. Steele, C.M., Josephs, R.A.: Alcohol myopia: Its prized and dangerous effects. Am. Psycholog. 45(8), 921 (1990)
24. Harvey, A.J., Kneller, W., Campbell, A.C.: The elusive effects of alcohol intoxication on visual attention and eyewitness memory. Appl. Cogn. Psychol. 27(5), 617–624 (2013)
25. Harvey, A.J.: Some effects of alcohol and eye movements on cross-race face learning. Memory. 22(8), 1126–1138 (2014)
26. Marple-Horvat, D.E., Cooper, H.L., Gilbey, S.L., Watson, J.C., Mehta, N., Kaur-Mann, D., et al.: Alcohol badly affects eye movements linked to steering, providing for automatic in-car detection of drink driving. Neuropsychopharmacology 33(4), 849–858 (2008)
27. Ahlström, C., Zemblys, R., Finér, S., Kircher, K.: Alcohol impairs driver attention and prevents compensatory strategies. Accid. Anal. Prev. 184, 107010 (2023)
28. Hume, R., Weyers, E.: Relationship between total body water and surface area in normal and obese subjects. J. Clin. Pathol. 24(3), 234–238 (1971)
29. Sayer, I., Parry, D., Barker, J.: Traffic calming-An assessment of selected on-road chicane schemes. TRL Report 313, Transport Research Laboratory, Crowthorne, Berkshire (1998)
30. Åkerstedt, T., Gillberg, M.: Subjective and objective sleepiness in the active individual. Int. J. Neurosci. 52(1-2), 29–37 (1990)
31. Smart Eye AB: Smart Eye Pro Manual, revision 10(1) (2023)
32. Zemblys, R., Niehorster, D.C., Komogortsev, O., Holmqvist, K.: Using machine learning to detect events in eye-tracking data. Behav. Res. Meth. 50(1), 160–181 (2018)
33. Zemblys, R., Niehorster, D.C., Holmqvist, K.: gazeNet: End-to-end eye-movement event detection with deep neural networks. Behav. Res. Meth. 51(2), 840–864 (2019)
34. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press, Oxford (2011)
35. Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., Nyström, M.: One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. Behav. Res. Meth. 49(2), 616–637 (2017)
36. Ahlström, C., Victor, T., Wege, C., Steinmetz, E.: Processing of eye/head-tracking data in large-scale naturalistic driving data sets. IEEE Trans. Intell. Transp. Syst. 13(2), 553–564 (2011)
37. Saez de Urabain, IR, Johnson, MH, Smith, TJ: GraFIX: A semiautomatic approach for parsing low-and high-quality eye-tracking data. Behav. Res. Meth. 47(1), 53–72 (2015)
38. Hessels, R.S., Niehorster, D.C., Kemner, C., Hooge, I.T.: Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). Behav. Res. Meth. 49(5), 1802–1823 (2017)
39. Tukey, J.W.: Exploratory Data Analysis. vol. 2. Addison-Wesley Publishing Company, Reading, MA (1977)
40. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2022). https://www.R-project.org/
41. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1), 1–48 (2015). https://www.jstatsoft.org/index.php/jss/article/view/v067i01
42. Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B.: lmerTest package: Tests in linear mixed effects models. J. Stat. Softw. 82(13), 1–26 (2017). https://www.jstatsoft.org/index.php/jss/article/view/v082i13
43. Hooge, I.T., Niehorster, D.C., Nyström, M., Andersson, R., Hessels, R.S.: Is human classification by experienced untrained observers a gold standard in fixation detection? Behav. Res. Meth. 50, 1864–1881 (2018)
44. Wong, J.T., Huang, S.H.: Attention allocation patterns in naturalistic driving. Accid. Anal. Prev. 58, 140–147 (2013)
45. Gaspar, J., Carney, C.: The effect of partial automation on driver attention: A naturalistic driving study. Hum. Factors 61(8), 1261–1276 (2019)
46. Crossland, M.D., Rubin, G.S., et al.: The use of an infrared eye-tracker to measure fixation stability. Optomet. Vis. Sci. 79(11), 735–739 (2002)
47. Blignaut, P., Beelders, T.: The precision of eye-trackers: a case for a new measure. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 289–292. ACM, New York (2012)
48. Moskowitz, H., Robinson, C., et al.: Effects of Low Doses of Alcohol on Driving-Related Skills: A Review of the Evidence. National Highway Traffic Safety Administration, Washington DC (1988)
49. Tivesten, E., Broo, V., Aust, M.L.: The influence of alcohol and automation on drivers' visual behavior during test track driving. Transport. Res. Part F: Traff. Psychol. Behav. 95, 215–227 (2023)
50. Caffier, P.P., Erdmann, U., Ullsperger, P.: Experimental evaluation of eye-blink parameters as a drowsiness measure. Eur. J. Appl. Physiol. 89, 319–325 (2003)
51. Åkerstedt, T., Anund, A., Axelsson, J., Kecklund, G.: Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. J. Sleep Res. 23(3), 242–254 (2014)
52. Ahlström, C., Zemblys, R., Jansson, H., Forsberg, C., Karlsson, J., Anund, A.: Effects of partially automated driving on the development of driver sleepiness. Accid. Anal. Prev. 153, 106058 (2021)
53. Martin, T.L., Solbeck, P.A., Mayers, D.J., Langille, R.M., Buczek, Y., Pelletier, M.R.: A review of alcohol-impaired driving: The role of blood alcohol concentration and complexity of the driving task. J. Forensic Sci. 58(5), 1238–1250 (2013)
54. Dunn, M.J., Alexander, R.G., Amiebenomo, O.M., Arblaster, G., Atan, D., Erichsen, J.T., et al.: Minimal reporting guideline for research involving eye tracking. Behav. Res. Meth. 1–7 (2023)
55. Holmqvist, K, Nyström, M, Mulvey, F.: Eye tracker data quality: what it is and how to measure it. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 45–52. ACM, New York (2012)

56. Müller, A.I., Weinbeer, V., Bengler, K.: Using the wizard of Oz paradigm to prototype automated vehicles: methodological challenges. In: Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications: Adjunct proceedings, pp. 181–186. ACM, New York (2019)

How to cite this article: Zemblys, R., Ahlström, C., Kircher, K., Finér, S.: Practical aspects of measuring camera-based indicators of alcohol intoxication in manual and automated driving. IET Intell. Transp. Syst. 1–20 (2024). https://doi.org/10.1049/itr2.12520

APPENDIX A: TEST TRACK LAYOUT

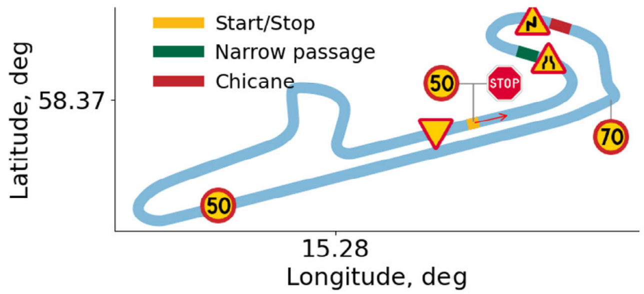

FIGURE A1 Mantorp Park (Sweden) test track layout used in our study. Red arrow shows driving direction. Because of logistical reasons *"Stop"* sign was only present in the first week of recordings.

APPENDIX B: ANALYSIS OF BRAC MEASUREMENT ORDER

In the analysis below, as fixed effects we entered BrAC measurement order and intoxication condition (with interaction term) and as random effects we had random intercepts for participants as well as by-participant random slopes for the effects of the measurement order and the intoxication condition; More specifically, we used the following formula: $lmer(BrAC \sim order * intoxication.condition + (1 + order + intoxication.condition|driverid)$; 0.2‰ intoxication condition was used as reference and for measurement order we used *Successive Differences Contrast Coding*.

TABLE B1 Linear mixed-effect regression model for BrAC measurements. Significant results are in bold.

	BrAC			
Predictors	**Estimates**	**CI**	**Statistic**	***p***
(Intercept)	0.19***	0.17 to 0.22	14.92	**1.72e-16**
AfterTrial1-BeforeTrial1	−0.04**	−0.08 to −0.01	−2.66	**8.92e-03**
BeforeTrial2-AfterTrial1	0.00	−0.03 to 0.03	0.21	8.30e-01
AfterTrial2-BeforeTrial2	−0.02	−0.04 to 0.01	−1.35	1.79e-01
0.5‰	0.34***	0.30 to 0.38	17.05	**4.02e-18**
0.8‰	0.59***	0.56 to 0.62	37.67	**3.06e-28**
1.0‰	0.76***	0.74 to 0.79	56.23	**1.42e-35**
AfterTrial1-BeforeTrial1:0.5‰	0.07***	0.04 to 0.11	3.94	**1.02e-04**
BeforeTrial2-AfterTrial1:0.5‰	−0.05*	−0.08 to −0.01	−2.53	**1.19e-02**
AfterTrial2-BeforeTrial2:0.5‰	0.02	−0.02 to 0.05	1.02	3.10e-01
AfterTrial1-BeforeTrial1:0.8‰	0.04*	0.01 to 0.08	2.44	**1.53e-02**
BeforeTrial2-AfterTrial1:0.8‰	−0.04	−0.07 to 0.00	−1.94	5.38e-02
AfterTrial2-BeforeTrial2:0.8‰	0.03	−0.00 to 0.06	1.72	8.71e-02
AfterTrial1-BeforeTrial1:1.0‰	0.04*	0.01 to 0.08	2.33	**2.06e-02**
BeforeTrial2-AfterTrial1:1.0‰	−0.06**	−0.10 to −0.02	−3.27	**1.19e-03**
AfterTrial2-BeforeTrial2:1.0‰	0.02	−0.01 to 0.05	1.19	2.34e-01
ICC				0.78
N drivers				35
Observations				490
Marginal R^2/conditional R^2				0.896/0.977

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

APPENDIX C: EYE TRACKING DATA QUALITY

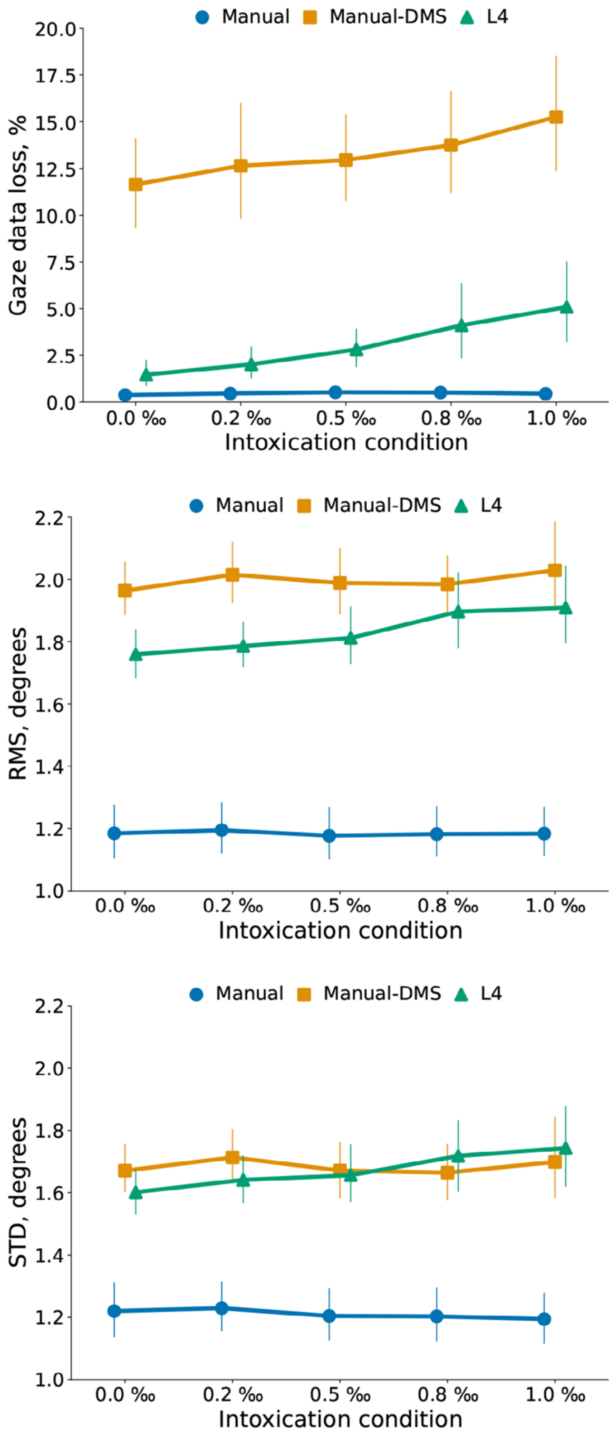

FIGURE C1 Eye tracking data quality measures (data loss and precision measured as root mean square of the difference between successive gaze position samples and the standard deviation of samples, respectively, RMS and STD) for different intoxication conditions with different camera setups.

APPENDIX D: AVERAGE SLEEPINESS LEVEL THROUGHOUT THE DRIVES

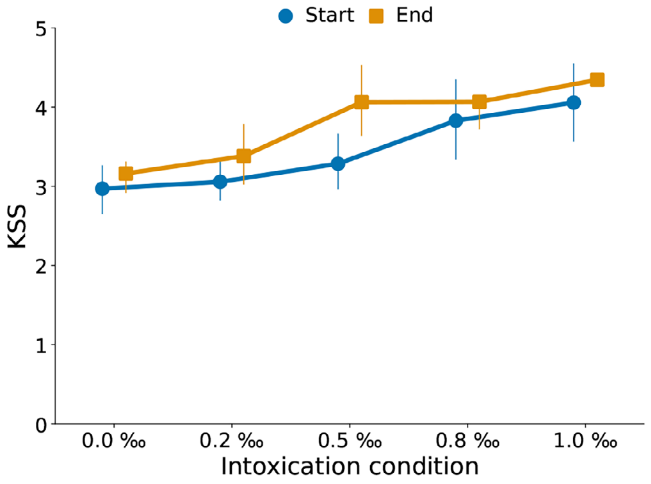

FIGURE D1 Averages of self reported sleepiness scores on *Karolinska Sleepiness Scale*[KSS, 30] across intoxication levels. "Start" refers to the score given before starting the trial (either driving on the test track or in simulator, depending on which the participant started with), and "End" refers to the score given after the last drive in that intoxication condition.