10th International Conference
on Driver Distraction and Inattention



DDI2026
GOTHENBURG
October 20-22

ABSTRACT BOOK

City of Gothenburg
VÄSTRA GÖTALANDSREGIONEN
smart eye
Fraunhofer IOSB

INDEX

Session 1 - HMI design and evaluation

Karsten Klaffer.....4
Designing Low-Distraction Vehicle Controls: Modality Selection and Interface Design Aspects

Isabel Mötsch and Natalia Sevcenko.....7
Are you still searching or already scrolling? – Layout-dependent effects of visual information quantities in in-vehicle information systems

Barbara Metz.....10
RealCarSIM: A new approach for the user testing of in-vehicle applications

Angus McKerral.....13
Counting the Cost: Validating Discrete Driver-IVIS Interaction Count as a Practical, Equipment-Free Index of In-Vehicle Infotainment System Distraction

Session 2 - Mind wandering, event boundaries and residual attention

Mark Vollrath.....17
Please talk to me - a systematic review of secondary tasks while driving

Avinoam Borowsky.....21
Driver distraction at roadway event boundaries: a simulator investigation within an event segmentation framework

Séverine Brelaud.....26
Hands on the wheel, head in the clouds? Mind-Wandering Signatures in Simulated Manual Driving

Yannick Lavril.....30
The impact of non-driving related task engagement on the attentional processing of a novel unexpected auditory information

Session 3 - Visual attention beyond fixation

Victor Meulenkamp.....34
Beyond Fixation: A Systematic Review of Peripheral Vision Use in Driving

Tibor Petzoldt.....40
Dynamic Advertising on Vehicles: Effects on Driver Glance Behavior

Katarina Pavic.....43
Reducing inattentional blindness to warning signals: the influence of meaning, emotional value, and task context

Malin Svärd.....46
Braking without looking: understanding driver responses to sudden lead-vehicle deceleration using a forced peripheral vision paradigm

Session 4 - Advances in driver state monitoring

Iris Jestin.....49
”What Does the Car Know?” Mapping Mental Models of Driver Monitoring Systems

Paul Hemeren.....53
I2Connect: Intelligent, Interactive, and Connected - Next Generation of Real-time Driver Safety Support Systems

Michele Guagnano.....56
Drowsiness and sleep prediction via imaging PPG: a multiscenario validation from Simulated to Naturalistic Driving

Poster briefings (Group A)

Nadja Schömig.....61
A Methodological Approach to Evaluate Driver Distraction During In-vehicle Interactions with Voice-based Large Language Models

Linda Pipkorn.....66
Hands On, Eyes On! Driver Monitoring and Behaviors in Hands-Free Level 2 Automation

Kajsa Weibull.....71
Reducing Ambiguity in Emergency Vehicle Interaction: Strengthening Communication with LED Displays

Jordanka Kovaceva.....74
Baseline gaze behavior of younger and older drivers under undistracted conditions in cyclist intersection conflicts

Aimé Tiadi.....76
Characterizing and Detecting the Driver Intended Head Movements During attentional and inattentional behaviours: An Open Road and Test Track Study

Alessandra Lioi.....84
Adaptive Automation of Human Information Processing in Driver-Road Interaction: A Driving Simulation Study in Motorway Environments

Jumpei Yamashita.....87
Where Luminance Guides Driving: Route-Specific Attention Revealed by Vehicle Control and Pupil Responses

Juraj Leonard Vertlberg.....90
Driver Distraction and Mobile Device Usage: Safety Performance Indicator Results for Croatia

Rita Helen Aarvold.....93
Enhancing driver attention through behavior change interventions: Insights from the Norwegian driving challenge

Session 5: Automation transitions, readiness and disengagement

Katharina Kummerer.....96
Partial driving automation can impose additional demand on the driver without impairing driving performance in a car-following scenario

Rafael C. Gonçalves.....101
Validation of an evidence accumulation model for driver readiness in assisted driving using driver ocular metrics

Joonbum Lee.....104
How L3 Driving Automation Will Actually Be Used: A Phase-Based Taxonomy of Usage Patterns and Design Considerations

Joonbum Lee.....110
From Old Bottles to New: Reframing Driver Distraction in an Era of Vehicle Automation

Session 6: Driver monitoring, impairment and assessment

Adriano Palao.....114
Euro NCAP 2029 Driver Monitoring Assessment Framework for Mitigating Driver Inattention Risks

Huy Phan.....117
Exogenous Disruptions of Visual Attention and Gaze Abnormality

Christer Ahlström.....122
Disentangling alcohol and sleepiness effects on driver monitoring system metrics

Poster briefings (Group B)

Sanjana Hassan Ananda Kumar.....126
Micro- and Macro-Steering Reversal Analysis for Personalized Drowsiness Detection

Claus Marberger.....130
Driver attention during partially automated driving: Insights from reference drivers' glance behavior

Mahdi Davari.....133
Beyond the Source of Distraction: Shared Neural Signatures of Internal and External Distraction in Drivers

Denny Kopp.....137
Calibrating Non-Invasive Impairment Detection Without Intoxication Ground Truth: An Indirect Application Approach

Karamjeet Kaur Sarna.....141
Driving with future intentions: examining the influence of prospective memory representations on driving performance

Rachael A. Wynne.....146
Seeing Through Occlusion: Validation of the Occlusion Test Protocol for IVIS Interactions

Victor Fabricius.....149
Comparing a camera monitor system (CMS) with conventional mirrors in heavy goods vehicles: Results from a reversing experiment

John Cliburn.....152
Comparing Eyes-Off-Road Behavior Between Tactile and Touchscreen Secondary Task Interfaces

Johan Karlsson.....155
Prerequisites for Adaptive ADAS: Learning from Fatal Crashes in Sweden

Session 7: Real-world interaction behaviour, conversational load and crash risk

Minxiang Zhao.....158
Conversational cognitive load meets lateral hazard exposure: How environment shapes crash risk

Allegra Ayala.....162
How Drivers Allocate Visual Attention During Extended Conversational AI Interactions

Mirta Zelenika Zeba.....166
Driver interaction with central display in real-world driving conditions

Session 8: Distraction and driver characteristics

Mette Møller.....169
Behaviour of cognitively distracted cyclists in T-junctions

Courtney M. Goodridge.....172
Age-Dependent Changes in Pupillometric Indicators of Cognitive Load During Hands-Off L2 Driving

Alexandria Noble.....176
Crash risk of eyes off road among teen drivers across age and driving experience

Session 9: Automation in use: engagement, gaze and user state

Yee Thung Lee.....179
Investigating drivers' visual scanning patterns under different levels of driving automation across daytime and after dark driving environments

Angus McKerral.....182
Anxiety as a Source of Inattention During Highly Automated Driving: Effects on Situation Awareness at Takeover

Sarah Schwindt-Drews.....185
Effects of Automated Driving Style and Visual Countermeasures on Motion Sickness During Visual NDRT

Joshua Domeyer.....189
Comparison of Driver In-Vehicle Technology Engagement for Hands-On and Hands-Free Partial Driving Automation

Session 10: Safety assessment, models and validation

Katja Kircher.....192
Evaluating attentional requirements at scale: A GIS-based proof of concept

Tuomo Kujala.....195
Possibility over probability: A worst-case approach to driver distraction and inattention assessment

Jonas Bärgman.....198
From glances to outcomes: How counterfactual risk estimates relate to real-world crash odds

Rafael C. Gonçalves.....202
Predicting Reaction Time Under L2 Assisted Driving and Cognitive Load: A Sensitivity Analysis of Eye-Tracking Features Using XGBoost and Explainable AI

Designing Low-Distraction Vehicle Controls: Modality Selection and Interface Design Aspects

Karsten Klaffer [1 *], Tina Gehlert [2], Mark Vollrath [1]

[1] Technische Universität Braunschweig, Department of Traffic and Engineering Psychology, Gaußstraße 23, 38106 Braunschweig
[2] German Insurers Accident Research (Unfallforschung der Versicherer, UDV), Wilhelmstraße 43 / 43G, 10117 Berlin
[*] karsten.klaffer@tu-braunschweig.de

Abstract: Modern vehicles are equipped with a large number of functions available to drivers. This is accompanied by a change towards touch- and speech-based interfaces while physical controls are being used less. However, to reduce driver distraction, it is crucial to use a suitable human-machine interface for operating a given function. Both the modality of the interface and task complexity affect driver distraction. Although this is targeted by Euro NCAP's General Vehicle Controls, it mostly focusses on simple single-step tasks and lacks information on the aspects that make each interface modality safe. Therefore, our study further investigated the effect of task complexity (simple or complex) and interface modality (physical controls, touch-based, and speech-based interfaces) as well as select design aspects for each modality (e.g. reachability, button sizes, error handling) on driver distraction. In a driving simulator experiment, participants (N = 130) operated a simple or complex function via generic interfaces. The results indicate that for a simple task there is no difference between modalities while for a complex task, speech-based interfaces lead to the least amount of distraction. For design aspects, the motor effort for physical controls and touchscreens should be kept minimal. Speech-based interfaces should use short commands and not switch modalities or introduce errors in the operation process.

1. Introduction

Modern vehicles are equipped with a wide range of functions that drivers can interact with, such as driving assistance, comfort, or infotainment. To provide function access, most manufacturers rely on a central touchscreen. As a result, the number of touchscreen-equipped vehicles is rising and screen sizes are growing (Bocksch, 2023; Stock, 2023). At the same time, physical controls (buttons, knobs) are phased out, and manufacturers offer voice-based interfaces as another way to operate functions.

1.1 Driver Distraction and Function Operation

It is well known that secondary tasks while driving may result in driver distraction and an increased crash risk (Ferdinand & Menachemi, 2014; Klauer et al., 2006). Given that function operation is the most frequently performed secondary task while driving (Lansdown, 2012; Young et al., 2019), it is important to minimize its impact on driver distraction.

Two elements are important in understanding distraction in the context of function operation: interface modality and task complexity. Both affect driver distraction: Voice-based interfaces usually result in the least amount of distraction, followed by physical controls, and touchscreens resulting in the highest distraction (e.g. Ma et al., 2022). At the same time, distraction increases with task complexity (e.g. Strayer et al., 2019; Zhong et al., 2024).

1.2 Minimizing Distraction by Design

In theory, vehicle manufacturers could provide function access via any interface modality, irrespective of task complexity. However, this may result in a function operation that is unsafe for driving. As a result, Euro NCAP included the design of the human-machine interfaces used for function operation in their 2026 rating scheme (Euro NCAP, 2025). The General Vehicle Controls assessment includes a list of both driving and non-driving related functions. Based on the use context (e.g. time criticality) of the function, four acceptable function implementations are derived: direct physical input, direct touch input, direct voice input, and menu-based touch input.

While the protocol is a welcome development regarding safe function implementation, it could be improved in two ways. First, it should consider more complex tasks with more than one operation step (e.g. adjusting an active navigation). Currently, these are only somewhat represented in the menu-based touch input with a maximum of two steps. Secondly, for each interface modality, more information on what aspects make it less prone to distraction would be useful. Designers would benefit from knowing how buttons should be designed or how the visual interface of the touchscreen should be organized, for example.

To this end, this study investigated how task complexity and interface modality affect driver distraction when operating in-vehicle functions in a driving simulator experiment. Additionally, it explored which specific design aspects are relevant for reducing driver distraction for physical controls, touch-, and speech-based interfaces.

2. Method

2.1 Sample

The final sample consists of N = 130 drivers (55 female, 74 male, 1 not specified). They were aged M = 39.8 years (SD = 19.0 years).

2.2 Materials

2.2.1 Driving Simulation

The Lane Change Task (LCT; International Organization for Standardization, 2010) was used to quantify the function operation task's demand. The LCT incorporates a simplified driving simulation in which participants drive along a three-lane road and have to perform lane changes. Each trial consists of 18 lane changes and lasts 3 minutes. A mean deviation (MDev) from an ideal lane was computed as an indicator of driving performance.

2.2.2 Function Operation Tasks

Two function operation tasks with different levels of complexity were employed. First, a time-critical task that requires immediate operation because of external circumstances and which can be done in a single step. In our experiment, the external cause was the lane change of the LCT so that indicating in the according direction was the derived function task. Second, a non-time-critical task with multiple steps was used. In the experiment, this was represented by typing in and calling a phone number dictated by the experimenter.

2.2.3 Interfaces: Modalities and Design Aspects

For all three modalities (physical controls, touch-, and speech-based interfaces), a generic interface was designed. The software for all three interfaces ran on an Apple iPad Air to provide basic visual and auditory feedback. It was placed to the right of the steering wheel.

For physical controls, buttons on the steering wheel were used. The touch-based interface was provided on the Apple iPad. Finally, the speech-based interface used voice-recognition and provided voice feedback. The basic experimental setup is shown in Figure 1.



Figure 1: Driving simulator setup with generic function operation interfaces.

For each interface type, individual versions were designed to determine the effect of select interface design aspects. For example, for the touch-based interface, the button placement or the visual structure of the interface were altered. Similarly, for physical controls and the speech-based interfaces, their design was also varied. In total, 51 design aspects were tested (12 for physical controls, 23 for the touch-based interface, 16 for the speech-based interface). To derive the design aspects a mix of design guidelines (e.g. European Union, 2008; National Highway Traffic Safety Administration, 2013), norms (e.g. International Organization for Standardization, 2009, 2017), and standards (Society of Automotive Engineers, 2015) was used.

In the reference version of an interface, all design aspects were implemented in an acceptable form. In the version for the aspect test, the according design principle was violated and therefore unacceptable.

2.3 Design and Procedure

Before the experiment, participants trained both the LCT and function operation tasks individually and then in parallel. After training was completed, the experiment began.

Per LCT-Trial, participants interacted with either the simple or complex task in a reference interface or the version for the interface aspect test. Each participant experienced only four of the 51 design aspects to keep the experiment at a suitable length (90 minutes).

3. Results

For data analysis, driving performance in the baseline (no function operation) was subtracted for each trial in which function operation was tested for each participant. This was done to account for individual differences in baseline driving performance. Additionally, mixed models with random intercepts at the individual and group level were used to account for additional effects.

3.1 Task Complexity and Modality

Figure 2 displays the Mean Deviation for all modalities and operation tasks.

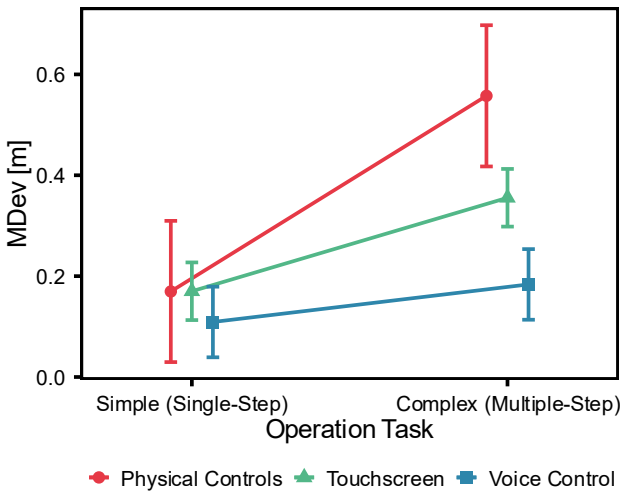


Figure 2: Mean Deviation (MDev) from ideal trajectory in the LCT for each combination of operation task and interface modality. Reference interfaces were used for function operation.

An ANOVA revealed significant main effects of modality and operation task on MDev, as well as an interaction effect (all $p < .05$). Pairwise comparisons showed no differences between modalities for the simple task. However, each pair of modalities differs in the complex task (all $p < .05$). Overall, the speech-based interface degraded driving performance the least.

3.2 Design Aspects

The design aspects are currently explored in more detail, however, some general conclusions can already be drawn.

For physical controls, the data indicates that they should be placed close to the driver and that interacting with the element should require little effort, i.e. no precise grabbing or movements required.

Similarly, touch-based interfaces should rely on simple inputs, i.e. tapping instead of sliding, and use large enough buttons for function operation. Additional interaction steps, for example for error handling, should be avoided.

Finally, speech-based interfaces should try to handle function operation with short commands and feedback. Also, operation should not switch modalities, e.g. by pressing a touch button because the voice control activated something wrong. Furthermore, it is important that the voice-based interface does not introduce errors in the operation process.

4. Discussion

While the Euro NCAP's General Vehicle Control assessment targets a safe function implementation, it only captures function operation in a simplistic manner. For improvement, more complex tasks with two or more operation steps should be considered. Furthermore, more interface modalities should be taken into account when looking at the complexity of a function. This is especially important for speech-accessible functions and their fallback option via physical controls or the touchscreen.

With an improved assessment, function operation could be further optimized towards safe function implementations.

References

Bocksch, R. (2023, September 4). *Infografik: Wie smart sind die Autos der Deutschen?* Statista. https://de.statista.com/infografik/30738/anteil-der-befragten-deren-auto-diese-funktionen-hat

Euro NCAP. (2025). *Euro NCAP Protocol—Safe Driving—Driver Engagement*. https://www.euroncap.com/en/for-engineers/protocols/2026-protocols/

European Union. (2008). *Commission Recommendation of 26 May 2008 on safe and efficient in-vehicle information and communication systems: Update of the European Statement of Principles on human-machine interface (notified under document number C(2008) 1742)*. http://data.europa.eu/eli/reco/2008/653/oj/eng

Ferdinand, A. O., & Menachemi, N. (2014). Associations Between Driving Performance and Engaging in Secondary Tasks: A Systematic Review. *American Journal of Public Health*, *104*(3), e39–e48. https://doi.org/10.2105/ajph.2013.301750

International Organization for Standardization. (2009). *Road vehicles Location of hand controls, indicators and tell-tales in motor vehicles* (ISO 4040:2009; Version 5). https://www.iso.org/standard/44856.html

International Organization for Standardization. (2010). *Road vehicles—Ergonomic aspects of transport information and control systems—Simulated lane change test to assess in-vehicle secondary task demand* (ISO 26022:2010). https://www.iso.org/standard/43361.html

International Organization for Standardization. (2017). *Road vehicles—Ergonomic aspects of transportation and control systems—Dialogue management principles and compliance procedures* (ISO 15005:2017). https://www.iso.org/standard/69238.html

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., Ramsey, D. J., & Virginia Tech Transportation Institute. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data* (DOT HS 810 594). https://doi.org/10.21949/1530253

Lansdown, T. C. (2012). Individual differences and propensity to engage with in-vehicle distractions – A self-report survey. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(1), 1–8. https://doi.org/10.1016/j.trf.2011.09.001

Ma, J., Li, J., & Gong, Z. (2022). Evaluation of driver distraction from in-vehicle information systems: A simulator study of interaction modes and secondary tasks classes on eight production cars. *International Journal of Industrial Ergonomics*, *92*, 103380. https://doi.org/10.1016/j.ergon.2022.103380

National Highway Traffic Safety Administration. (2013). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices* (No. 81). https://www.federalregister.gov/documents/2013/04/26/2013-09883/visual-manual-nhtsa-driver-distraction-guidelines-for-in-vehicle-electronic-devices

Society of Automotive Engineers. (2015). *Guidelines for Speech Input and Audible Output in a Driver Vehicle Interface* (SAE J2988). https://doi.org/10.4271/J2988_201506

Stock, K. (2023, February 13). *Every car has a giant touch screen now* [The Detroit News]. Bloomberg. https://www.detroitnews.com/story/business/autos/2023/02/13/every-car-has-agiant-touch-screen-now/69898312007/

Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D. J., & Biondi, F. (2019). Assessing the visual and cognitive demands of in-vehicle information systems. *Cognitive Research: Principles and Implications*, *4*(1). https://doi.org/10.1186/s41235-019-0166-3

Young, K., Osborne, R., Koppel, S., Charlton, J., Grzebieta, R., Williamson, A., Haworth, N., Woolley, J., & Senserrick, T. (2019). What are Australian drivers doing behind the wheel? An overview of secondary task data from the Australian Naturalistic Driving Study. *Journal of the Australasian College of Road Safety*, *30*(1), 27–33. https://doi.org/10.33492/jacrs-d-18-00085

Zhong, Q., Zhi, J., Xu, Y., Gao, P., & Feng, S. (2024). Assessing driver distraction from in-vehicle information system: An on-road study exploring the effects of input modalities and secondary task types. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-71226-4

Layout-dependent effects of visual information quantities in in-vehicle information systems

Isabel Mötsch [1], Natalia Sevcenko [2], Ansgar Gerlicher [3]

[1] Stuttgart Media University, Stuttgart, Germany
[2] Mercedes-Benz AG, Research & Development, Sindelfingen, Germany
[3] Stuttgart Media University, Stuttgart, Germany

Abstract: This study examines the distraction potential of labeled icons in In-Vehicle Information Systems (IVIS) depending on the layout (grid vs. scroll), information quantity (12, 15, 18, 20 items), and their interaction. In a driving simulator study (two-factor within-subjects design, n = 30), driver distraction was evaluated during a visual search task based on glance and driving behavior as well as task duration. For both layouts, total glance time and frequency towards IVIS, speed and distance variability, as well as task duration increased significantly with the number of items. This increase was most pronounced between 15 and 18 items. Interaction effects found on total glance time and task duration confirmed the influence of layout on the effect of the amount of visual information. While the number of items had no significant influence in the scroll layout, the values in the grid layout increased. These results suggest that the grid layout is more advantageous when displaying up to 15 items. For more than 15 items, the scroll layout indicates less distraction. Interestingly, the standard deviation of lateral position (SDLP) was lower across all quantities in the grid layout. Additional analyses revealed advantages for targets which are located on the first scroll page and known to the drivers. Implications for a quantity-dependent choice of layout for safer IVIS are derived.

1. Introduction

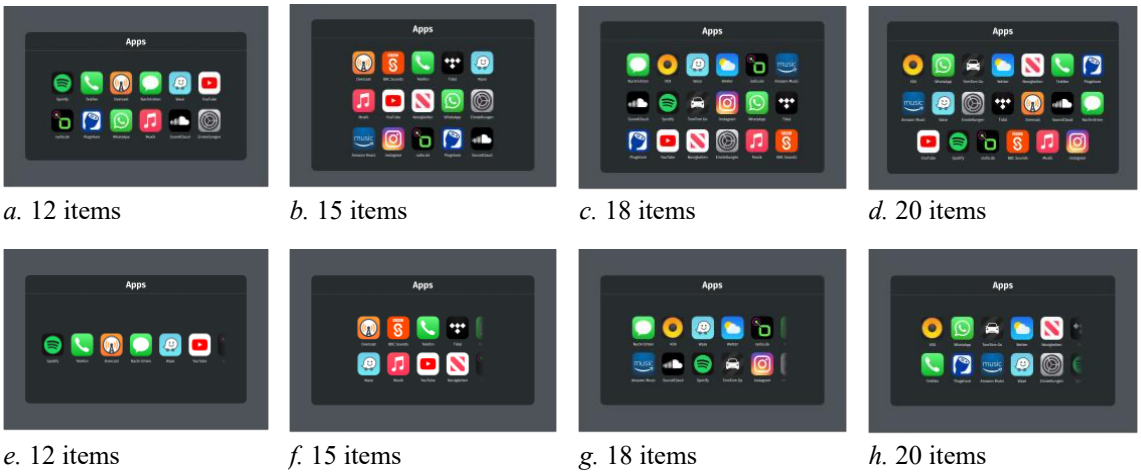
Increasingly larger formats and higher resolutions in In-Vehicle Information Systems (IVIS) allow more information to be displayed simultaneously. At the same time, driver distraction is considered one of the major risk factors for road safety, accounting for 17.8 % of fatal crashes in Canada in 2023 (Transport Canada, 2025). Displaying more visual information on an IVIS tends to impair the efficiency of visual search (Feng et al., 2018; Wolfe, 2007) while increasing the in-vehicle glance duration (Kujala & Saariluoma, 2011). In addition, the layout affects how information is visually perceived and processed (Li et al., 2017; Patten, 2007). This study examines the effects of information quantity, interface layout and their interaction on driver distraction.

2. Method

2.1 Study Design

Visual information items are defined as graphics that distinguishes themself visually from other items and present information. In this study they are exemplified by app icons, as these are used across various systems and allow the examination of both familiar and unfamiliar visual elements. To answer the research questions, a driving simulator study was conducted using a within-subjects design (n = 30) with the factors layout (scroll, grid) and quantity (12, 15, 18, 20 items) (Fig. 1).

The study was conducted in a static driving simulator as shown in Fig. 2. A Windows Surface Pro tablet was used for presenting the secondary tasks. Glance behavior was captured via Dikablis Glasses 3 eye-tracker (Ergoneers, n.d.).



a. 12 items *b.* 15 items *c.* 18 items *d.* 20 items

e. 12 items *f.* 15 items *g.* 18 items *h.* 20 items

Fig. 1. *Examined information quantities in* ***(a)-(d)*** *grid layout and* ***(e)-(h)*** *scroll layout (page 1 of 2).*

Driving data was recorded with the simulator. In addition, a detection-response task (DRT) was used to measure the reaction time to sudden events in road traffic (International Organization for Standardization [ISO], 2016).



***Fig. 2**: Driving Simulator Setup with DRT and visual search task.*

2.2 Tasks

The driving task was a simple car-following scenario, as described in the Alliance of Automobile Manufacturers guideline (2006). The subjects were instructed to maintain 50 m from the leading vehicle, which was traveling at a constant speed of 80 km/h. In addition, they responded to the DRT and performed a visual search task, which consisted of locating and selecting specific apps. The participants were instructed to prioritize the driving task and to perform the secondary tasks only when doing so would not impair their driving performance.

2.3 Analysis Strategy

The evaluated glance metrics included averaged values of single glance duration [ms], total glance time [s], and number of glances per task. The DRT captured reaction time and missed stimuli. Driving performance was quantified using standard deviation of the distance to the leading vehicle [m] (SDD), speed [km/h] (SDS), and lateral lane position [m] (SDLP). In addition, the task duration was assessed.

The data were analyzed using linear mixed models in jamovi (GAMLj3). In the case of significant main effects or interactions ($\alpha = 0.05$), post-hoc comparisons with Tukey correction were conducted to determine the differences between the levels. The error bars in all plots indicate 95 % confidence intervals.

3. Results

3.1 Layout Effect

Only the SDLP showed a significant layout effect across all quantities (Fig. 3). Lane keeping was more stable in the grid layout than in the scroll layout ($p < .001$, $R^2_{part} = .027$). The effect size indicates a small effect.

3.2 Quantity Effect

A significant quantity effect across both layouts was observed for total glance time ($p < .001$, $R^2_{part} = .055$), number of glances ($p = .001$, $R^2_{part} = .039$), SDS ($p = .006$, $R^2_{part} = .028$), SDD ($p = .001$, $R^2_{part} = .035$), and task duration ($p = .003$, $R^2_{part} = .031$). In the pairwise comparisons, significant differences were only found between 15 and 18 items, indicating an increase in visual-perceptual load between these quantities (Fig. 4).

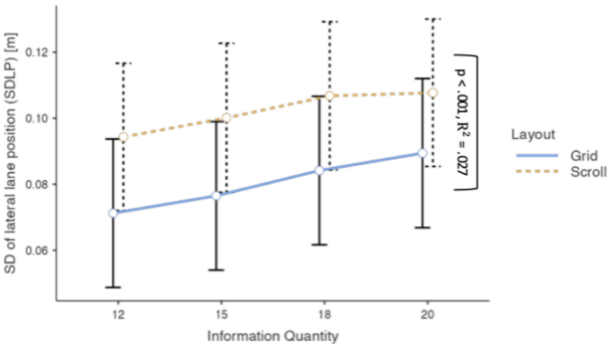


***Fig. 3:** Layout effect across all quantities on SDLP.*

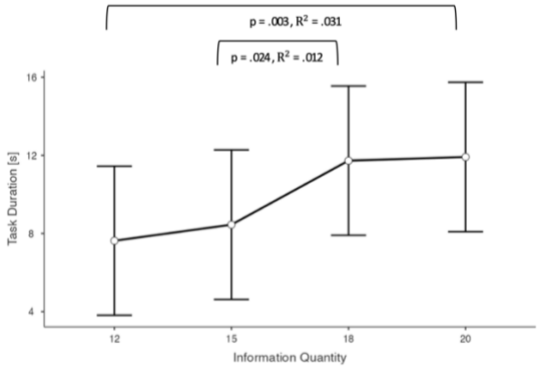


***Fig. 4**: Quantity effect across both layouts on task duration.*

3.3 Interaction Effect

It was assumed that the influence of the information quantity depends on the layout. Such an interaction effect was found for the total glance time ($p = .038$, $R^2_{part} = .023$) and task duration ($p = .031$, $R^2_{part} = .020$). Figure 5 shows the different increases in task duration between the quantities depending on the layout. The strongest increase was between 15 and 18 items in the grid layout, while in the scroll layout the task duration remained the same across all quantities.

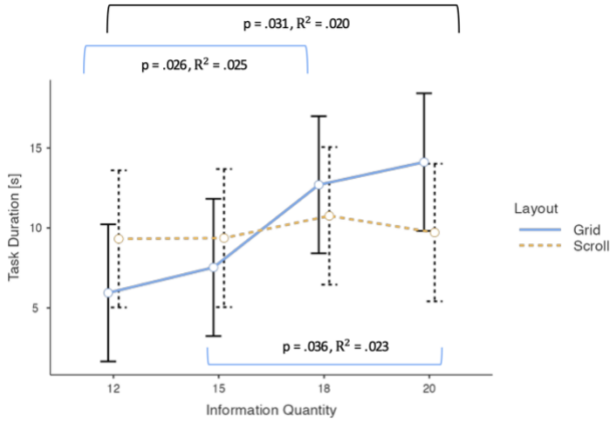


***Fig. 5:** Interaction effect (layout x quantity) on task duration.*

3.4 Exploratory Findings

The familiarity of the target app (according to a self-reported familiarity scale) had a significant impact on all variables examined except for DRT missing rate. With small (and medium for total glance time) effect sizes, glance behavior, driving performance, and task duration showed impairment at low familiarity levels.

The position of the target app in the scroll layout (first vs. second page) showed a significant effect for five of the nine variables examined. The total glance time, number of glances, SDD, SDLP and task duration were significantly higher when the target app was located on the second page instead of the first, and a scroll interaction became required.

4. Discussion

The quantity effects correspond to the literature, which reports lower search efficiency and longer task duration with an increasing number of items (Feng et al., 2018; Halverson & Hornof, 2004; Wolfe, 2007). Significant increases in the pairwise comparisons between 15 and 18 items indicates that additional visual observation might be necessary for perceiving 18 items. This is supported by the higher number of glances and longer total glance duration with the same average single glance duration.

Interaction effects show that the layout influences how the information quantity affects the driver distraction. They indicate that the dependent variables are more sensitive to an increase in the visual information quantity in the grid layout, while those in the scroll layout are less sensitive. Figure 5 shows a switch of the more advantageous layout. While the grid layout appears to be less distracting than the scroll layout for a smaller number of items (12 and 15), the total glance time and task duration increase for 18 items in the grid layout, exceeding the previously higher scroll values. At this information quantity, a grid layout might therefore be considered more distracting than a scroll layout.

Layout-related differences in SDLP point to greater lateral instability with the scroll layout than the grid layout by removing the hand from the steering wheel and performing a swiping hand movement.

As only app icons with a uniform design were used as stimuli, the findings cannot be applied to text-dominant or strongly different items. Factors such as colour, visual complexity, style, contrast, and placement were not systematically examined and should be addressed in future research.

4.1 Implications for the design of IVIS

Our findings indicate potential implications for the design of IVIS with visual items. The layout may be chosen depending on the intended number of items to be displayed. For smaller quantities of up to 15 items, a single-page grid layout without scroll interaction might be beneficial. When more than 15 items are displayed, it may be advantageous to divide the content across two pages rather than extending the grid. However, if the number of items exceeds 20 or if more than one scroll interaction is required, the distraction effects need to be re-examined. To support efficient search, the familiarity of items may be considered, and frequently used items might be placed on the first page of a scroll layout.

5. Conclusions

This study provides further empirical insights into minimizing driver distraction potential caused by visual items on the IVIS. It shows that distraction depends on the layout, information quantity and their interaction and offers practical recommendations for optimizing IVIS design.

References

Alliance of Automobile Manufacturers. (2006). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In- Vehicle Information and Communication Systems.

Transport Canada. (2025). *Canadian motor vehicle traffic collision statistics: 2023*. Government of Canada. https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics/2023/canadian-motor-vehicle-traffic-collision-statistics-2023

Ergoneers. (n.d.). Dikablis Glasses 3: Quick Start Guide [Technical documentation]. https://ergoneers.com/faq-downloads/manuals/Dikablis-Glasses-3/QS-Guide-DG3_Cable-and-Wireless.pdf

Feng, F., Liu, Y., & Chen, Y. (2018). Effects of Quantity and Size of Buttons of In-Vehicle Touch Screen on Drivers' Eye Glance Behavior. International Journal of Human–Computer Interaction, 34(12), 1105–1118. https://doi.org/10.1080/10447318.2017.1415688

Halverson, T., & Hornof, A. J. (2004). Local Density Guides Visual Search: Sparse Groups are First and Faster. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(16), 1860–1864. https://doi.org/10.1177/154193120404801615

International Organization for Standardization. (2016). ISO 17488: Road vehicles — Detection-response task. https://www.iso.org/standard/59887.html

Kujala, T., & Saariluoma, P. (2011). Effects of menu structure and touch screen scrolling style on the variability of glance durations during in-vehicle visual search tasks. Ergonomics, 54(8), 716–732. https://doi.org/10.1080/00140139.2011.592601

Li, R., Chen, Y. V., Sha, C., & Lu, Z. (2017). Effects of interface layout on the usability of In-Vehicle Information Systems and driving safety. Displays, 49, 124–132. https://doi.org/10.1016/j.displa.2017.07.008

Patten, C. J. D. (2007). Cognitive workload and the driver (Doctoral dissertation, Department of Psychology, Stockholm University). https://www.diva-portal.org/smash/get/diva2:196947/FULLTEXT01.pdf

Wolfe, J. M. (2007). Guided Search 4.0. In W. D. Gray (Ed.), Integrated Models of Cognitive Systems (pp. 99–119). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195189193.003.0008

RealCarSIM: A new approach for the user testing of in-vehicle applications

Barbara Metz*, Marcus Schmitz, Nora Merkel, Christian Mark, Alexandra Neukum

WIVW GmbH, Robert-Bosch-Str. 4, 97209 Veitshöchheim, Germany
* barbara.metz@wivw.de

Abstract: With emerging new desing options for in-cabin applications as well as driver monitoring systems becoming mandatory in the EU, the user centered testing of in-cabin applications for different vehicle models becomes more and more important. To avoid extensive testing on test tracks, an approach is needed which allows the easy integration of changing vehicle models in a driving simulation. As a solution, a simulation approach called RealCarSIM has been developed and implemented. In this approach, the vehicle under test is physically integrated into a driving simulator and it is interfaced with the simulation infrastructure to serve as input device for driving (steering angle, pedal positions) but also to put defined in-cabin applications (warning sounds, advanced driver distraction warning) into working mode. By this, RealCarSIM allows the flexible, controlled and repeated testing with normal drivers in varying driving environments (highway, urban, rural, test track), during varying driving tasks and varying traffic scenarios like any other driving simulator, but with the vehicle under test serving as simulator mockup. The RealCarSIM approach is presented for three different applications: UX evaluation of adaptive HMI solutions, evaluation of acceptability of different warning sounds and testing of advanced driver distraction warning (ADDW) e.g. for vehicle homologation. The advantage of the RealCarSIM approach is that it allows in-cabin applications to be tested in the vehicle under test, but under the controlled and replicable conditions of a simulated environment.

1. Introduction

Currently, there is rapid development of in-vehicle in-cabin applications and driver monitoring systems. For example, new design options for in-vehicle displays, such as large, curved or programmable displays, are entering the market, and driver monitoring systems are becoming mandatory and are addressed by dedicated regulations, such as REGULATION (EU) 2023/2590 for ADDW homologation, REGULATION (EU) 2021/1341 for DDAW homologation and DCAS Regulation (UN) 2025/17 for L2+ systems. Consequently, standardized, user-centred tests that address UX issues or evaluate the performance of driver monitoring systems are becoming a mandatory requirement for the development and homologation of in-cabin applications.

Driving simulators are an effective way of testing in-cabin applications under controlled and replicable conditions with a sample of non-professional drivers, eliminating any safety risks. Compared to test tracks, driving simulators allow testing in a variety of driving environments, including challenging driving tasks and other road users. However, standard driving simulators come with a fixed vehicle mock-up that usually allows only limited design adaptation.

The RealCarSIM approach presented here combines the advantages of simulator-based testing with the ability to easily switch between different vehicle models. After few integration steps, the vehicle under test is an integral part of the simulator. The communication between the vehicle and the simulation environment is realized by a suitable interface, e.g. via a vehicle bus.

2. Method

2.1 Setup of RealCarSIM

To integrate a vehicle physically into the RealCarSIM, the vehicle under test is positioned in front of a projection system, which is used to display the simulated environment. Here different display solutions have already been used like a LED wall (application 2) or 138" rear projection screen with an ultra-short throw laser projector (application 1 and 3). The front wheels are positioned on turntables that allow the wheels to turn as a reaction to steering inputs while the vehicle is stationary. Displays are also attached to the vehicle to serve as rear and side mirrors, and if needed as the cluster display. For the sound simulation, a subwoofer is placed behind the driver's seat and satellite speakers are installed in the driver's and passenger's doors. Figure 1 shows pictures of an examplary RealCarSIM setup.

2.2 Connection of vehicle to RealCarSIM

The simulation works with the simulation software SILAB® 7.3. This software is developed for human factors studies and allows the flexible simulation of various road environments (e.g. urban, rural roads, highways), interaction with other road users (vehicles, bicycles, pedestrians) and driving tasks. In addition, it offers interfaces to connect various external hardware components to the simulation. The CAN interface of the software package is used to read relevant information from the vehicle from a CAN bus that is connected via a USB CAN card. The information from this CAN bus (e.g. steering angle, accelerator pedal position, brake pedal position, indicator status) is used as input parameters for the driving simulation and controls the simulated vehicle in the simulated environment

For advanced applications, where certain components of the vehicle hardware shall be integrated in the simulation in a way that allows real-time interaction like in real traffic, these vehicle components are connected separately to the simulation software.



Fig. 1: *Vehicle integrated into RealCarSIM environment with projection of the main view on a LED wall, display used as left mirror and wheel positioned on turntable.*

3. Examples for user tests with RealCarSIM

3.1 Evaluation of adaptive HMI solutions

The first example study with RealCarSIM examined driver state dependent adaptations of vehicle functions by a so-called digital assistant. To achieve this, a specific driver state (happy, relaxed / tired) was induced in the participants at the start of a cross-country drive through instructions and the design of the simulation environment (e.g. weather and time of day). Shortly after setting off, the digital assistant contacted the participants via speech output and suggested adapting six vehicle functions (interior temperature, interior ventilation, seat ventilation, seat massage, ambient lighting and media playback volume) based on their state. For this study, an Audi A6 eTron was integrated into the RealCarSIM environment (see Figure 2). Once the participants agreed with the suggestion, the functions were adjusted accordingly. The subjects then evaluated the digital assistant and the implementation of the state-induced adaptations. Forty drivers participated in the study.

The requirement for the study was that the participants experience the vehicle functions (e.g. seat massage) or the change in functions (e.g. increase in volume) in the vehicle. This included adjusting the function displays in the vehicle (e.g. an increase in interior temperature from 20° to 21° being displayed). The aim was for the user experience in the vehicle interior to be as realistic as possible. This is why the RealCarSIM was chosen as test setup for this study.



Fig. 2: *RealCarSIM setup with Audi A6 eTron.*

3.2 Evaluation of sound concepts for in-vehicle warning systems

Another study using RealCarSIM evaluated three sound concepts for in-vehicle warning systems such as park distance warning, lane departure warning, hands-on warning and front collision alert. The main evaluation criteria were how well each sound fits the system, how annoying each sound was, which sound was preferred for each system and which sound concept was preferred overall. The study was conducted with 32 drivers.

The study had two main requirements. Firstly, participants had to perceive the sounds in the original vehicle via the original sound system. Secondly, participants experienced the warning sounds in a realistic traffic scenario to evaluate how well each sound matched the criticality of the traffic situation in which it was triggered. To fulfil these requirements, a KIA EV9 GT series vehicle was integrated into the RealCarSIM (see Figure 3). The RealCarSIM provided ideal conditions, as there was no external background noise (e.g. from changing surrounding traffic or wind conditions) that could change between participants. All participants experienced the same noise conditions so that the results were achieved under comparable circumstances.



Fig. 3: *RealCarSIM setup with KIA EV9.*

3.3 Testing of ADDW systems

A prototypical nearproduction vehicle was used to test the potential of RealCarSIM for the evaluation of ADDW systems e.g. for vehicle homologation. The main requirement was that testing be conducted with the prototype vehicle, since the final performance of the ADDW depends on the final hardware and cabin design (e.g. the position of the driver camera and the position of the fixation points).

For the setup, the ADDW control unit was connected to the simulation via a CAN gateway. This allowed the ADDW to react to drivers' state and to interact with the simulated environment as it would in real road conditions. Warnings and other information from the ADDW system were transferred to the vehicle, resulting in the corresponding auditory and visual warnings.

The full experimental protocol for ADDW homologation according to REGULATION (EU) 2023/2590, including day and night testing, multiple fixation points, and different speed ranges, was transferred from the test track to the RealCarSIM using a simulated test track.

For three drivers, the test protocol was conducted in both environments, and the data were compared. This comparison revealed comparable ADDW behaviour not only with regard to potential homologation. Subtle differences in warning times observed on the test track between drivers and across different fixation points were also replicated in the RealCarSIM (see figure 4 for an example).

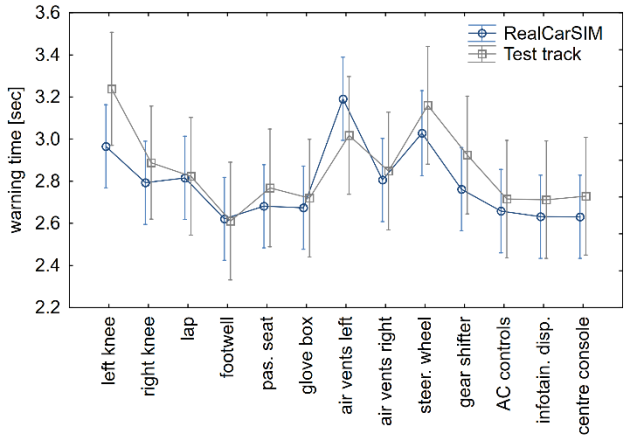


Fig. 4: *ADDW warning times for different fixation points on the test track and in the RealCarSIM.*

The experiment shows that comparable results can be achieved in the RealCarSIM while the test effort could be reduced significantly (no rental of a test track, no transfer of the test vehicle and participants to the test track etc.).

4. Discussion

The presented studies underline the potential of the RealCarSim approach through three different applications. It enables researchers to use different (near)production vehicles as mock-ups in a driving simulator. This allows mature in-cabin applications and up-to-date vehicle interiors to be evaluated in user tests, making use of the driving simulator's ability to repeatedly and realistically simulate different driving environments and tasks in reproducible and safe conditions. Currently, the main drawback of the RealCarSIM approach is that it requires access to, and detailed knowledge of, the vehicle structure in order to use vehicle information (e.g. from the CAN bus) as input parameters for the simulation software, and even more in-depth knowledge if certain in-car applications (e.g. ADDW) are to be included in the testing. The development of an additional set of sensors that can flexibly measure pedal position and steering angle would facilitate easy access to relevant input from the vehicle, simplyfying the integration of various series production vehicles into the RealCarSIM.

5. Conclusions

As in-cabin applications and driver monitoring systems face stricter regulations for series production, evaluating Human-Machine Interface (HMI) solutions has become essential to vehicle development. The RealCarSIM approach presented here offers an opportunity to move part of this process into a simulated environment, taking advantage of the ability to conduct repeated and standardized testing while using the vehicle under test with its final hardware and software components.

References

European Commision (2021). **Commission Delegated Regulation (EU) 2021/1341** of 23 April 2021 supplementing Regulation (EU) 2019/2144. https://eur-lex.europa.eu/eli/reg_del/2021/1341/oj/eng

European Commision (2023). **Commission Delegated Regulation (EU) 2023/2590** of 13 July 2023 supplements Regulation (EU) 2019/2144. https://eur-lex.europa.eu/eli/reg_del/2023/2590/oj/eng

United Nations (2024). **Addendum 170 – UN Regulation No. 171: Uniform Provisions Concerning the Approval of Vehicles with regard to Driver Control Assistance Systems.** https://unece.org/sites/default/files/2025-03/R171e.pdf

Counting the Cost: Validating Discrete Driver- In-Vehicle Infotainment System (IVIS) Interaction Count as a Practical, Equipment-Free Index of IVIS Distraction

Angus McKerral[1, 2,*], Jon-Paul Cavallaro[1], Rachael A. Wynne[1], Alexander Thorpe[1], Ami Eidels[1], Scott D. Brown[1], Guy E. Hawkins[1], Kristen Pammer[1,2]

[1] The University of Newcastle, University Drive, Callaghan, Australia
[2] The University of Melbourne, Grattan St, Parkville, Melbourne, Australia

[*]angus.mckerral@unimelb.edu.au

Abstract: Assessing the attentional costs of In-Vehicle Infotainment System (IVIS) interaction typically relies on established on-track and laboratory measures: the Detection Response Task (DRT), which indexes residual attentional capacity as a proxy for cognitive load, and the Visual Occlusion Test (VOT), which quantifies visual attentional demand. Both require specialised equipment, controlled testing environments, and trained personnel, restricting their use to well-resourced settings. This paper reports the development and validation of a simpler alternative: a count of the discrete driver inputs required to complete an IVIS task, referred to as step count. Step count was operationalised from video-coded observations across on-track and stationary test sessions involving a range of IVIS systems in the Australian market. Step counts were stable across assessment contexts (r > .67), and cross-validated mixed-effects models showed robust predictive relationships with all three criterion measures: on-track DRT miss rate (r = .71), VOT Total Shutter Open Time (r = .65), and simulated-driving tracking error (r = .81). Each additional step was associated with approximately a 1 percentage point increase in the probability of a DRT miss, and 270–800 ms of additional eyes-off-road time. A parsimonious, system-agnostic model specification was favoured over more complex alternatives, preserving the metric's applicability to novel IVIS designs. These findings establish step count as a valid, generalisable index of IVIS interaction demand, administrable by non-expert observers without instrumentation.

1. Introduction

Driver distraction attributable to In-Vehicle Infotainment Systems (IVIS) is a significant road safety concern. Drivers spend between 30% and 40% of their time behind the wheel engaged in secondary tasks (Metz et al., 2014), and modern IVIS intensify this risk by embedding complex, multi-step interaction sequences into the driving environment. Executing these interactions requires the driver to simultaneously redirect visual attention, engage cognitive resources, and produce manual inputs, sustaining attentional withdrawal from the forward driving scene across the full duration of each task (Strayer et al., 2011).

Research has established sensitive criterion measures of the cognitive and visual costs of IVIS interaction. The Detection Response Task (DRT; ISO 17488:2016) measures drivers' residual attentional capacity by requiring responses to a peripheral stimulus concurrent with a secondary task: elevated miss rates and slowed response times indicate that the secondary task is consuming attentional resources that would otherwise be available for hazard detection (Strayer et al., 2015). The standard tactile DRT configuration is selective for cognitive demand, though variants have been developed to extend its sensitivity to visual task demands as well (Strayer et al., 2016). The Visual Occlusion Test (VOT) provides a complementary index of the visual attentional requirements of an IVIS task by quantifying the minimum eyes-available time a driver needs to complete it (Total Shutter Open Time; TSOT). Together, these measures have supported rigorous comparative evaluation of IVIS systems across vehicle fleets (Strayer et al., 2017; Strayer et al., 2019). Both, however, require specialist equipment, controlled test environments, and trained personnel, limiting their practical reach.

The notion that objective task complexity metrics can predict driver workload has been previously explored in the literature. Getty et al. (2018) investigated menu depth – the hierarchical complexity of voice command sequences – as a predictor of driver demand and interaction time across production vehicles, finding that structural complexity of the voice interface was associated with increased workload. More broadly, temporal measures of task engagement are among the most sensitive discriminators of IVIS workload differences between vehicles, outperforming DRT-based and subjective measures in variance and power analyses (McDonnell et al., 2021). These findings are consistent with resource-competition accounts of driver distraction (Strayer et al., 2011), which predict attentional impairment in proportion to secondary-task depth, and with the broader principle that task complexity is central to mental workload (Wickens, 2008).

What has not been established is a formally cross-validated predictive model linking a multimodal step count metric – spanning touchscreen, physical control, and voice interactions – to criterion-validated measures of cognitive load and visual attentional demand. Getty et al. (2018) validated a structural measure of voice interaction complexity against composite demand and task-time outcomes only, in a voice-only context. The present paper addresses this gap by operationalising step count as a modality-agnostic input metric, establishing its cross-context stability, and evaluating its cross-validated predictive relationships with DRT, VOT,

and simulated-driving criterion measures. Cost functions are derived describing the quantitative attentional cost of each additional interaction.

2. Method

2.1 Participants and Assessment Contexts

Twenty participants completed a repeated-measures protocol across a range of IVIS designs available in the Australian market, including touchscreen-dominant systems, hybrid voice-and-touch interfaces, and integrated third-party platforms (Apple CarPlay, Android Auto). Testing was conducted at the Australian Automotive Research Centre (AARC), Victoria. Two criterion assessment contexts were used. In the on-track context, participants drove a 4.2 km closed highway circuit at 80 km/h while performing IVIS tasks and responding to a tactile DRT (ISO 17488:2016). In the stationary context, the same participants completed IVIS tasks in parked vehicles assessed via a Visual Occlusion Test (VOT) with visual occlusion goggles and a computer-based lateral-tracking task with embedded visual DRT. A standardised task battery spanning phone operations, navigation, media control, and messaging was used in both contexts (10 trials per task type per context).

2.2 Step Count Operationalisation

A step was defined as any discrete driver input to the IVIS: touchscreen contacts, physical button or dial activations, verbal commands, and verbal responses to system prompts. Failed interactions, where the system did not recognise a command and the driver was required to repeat the input, were counted as additional steps on the basis that they impose equivalent attentional demands on the driver regardless of outcome. Step counts were coded from post-hoc review of video recordings by trained observers. One participant was removed from VOT-dependent analyses (including the VOT step-count model) for failure to follow the VOT protocol, leaving n = 19 for those models; the full sample of 20 was retained for DRT and tracking task analyses. A preliminary cross-context stability analysis correlated IVIS-level step count estimates derived from on-track models with those from stationary models, testing whether systems requiring more interactions on-track also did so in stationary conditions.

2.3 Model Comparison

Three linear mixed-effects model specifications were compared for each criterion outcome, using participant-level mean step counts per IVIS task as the predictor: (1) a random-intercept-only model (Outcome ~ nSteps + (1 | Participant)); (2) a model adding IVIS-specific step slopes; and (3) a random-intercept-and-slope model. Model selection prioritised Leave-One-Out Cross-Validation (LOOCV) accuracy over in-sample fit, assessing how well the model generalises to IVIS systems not in the training data. All analyses were conducted in R (v. 4.3.2) using *lme4* (Bates et al., 2015).

3. Results

3.1 Cross-Context Stability of Step Count

IVIS-level step count estimates correlated substantially between on-track and stationary models ($r > .67$), confirming that systems requiring more interactions during on-track driving also required more during stationary testing. This cross-context consistency is a prerequisite for using step count as a context-independent index of interaction demand, and provides the empirical basis for its use as a predictor of criterion distraction measures.

3.2 Model Selection

The random-intercept-only specification was selected as the final model for both the DRT and VOT outcomes. For the DRT, a more complex random-slope model produced modestly higher LOOCV accuracy ($r = .75$ vs. $r = .71$), but required a participant-varying step coefficient that cannot be applied to a new vehicle without prior data from that vehicle. For the VOT, the simpler model matched the cross-validated performance of more complex alternatives exactly. A metric whose predictive relationship with distraction is fixed and system-agnostic is a design requirement for an instrument intended to generalise beyond the vehicles on which it was calibrated.

3.3 Predictive Validity and Cost Functions

Table 1 summarises cross-validated predictive accuracy across all criterion measures. The positive step-to-distraction relationship was consistent across three independent outcomes and both assessment contexts. Cost functions derived from the selected model indicated that each additional interaction step was associated with approximately a 1 percentage point increase in DRT miss probability (slope = 0.011) and approximately 270–800 ms of additional eyes-off-road time, with later steps accumulating disproportionately greater visual cost as task complexity grows.

Table 1. *Cross-validated predictive accuracy of step count against criterion measures of cognitive load and visual attentional demand.*

Criterion measure	Assessment context	Cross-validated r (LOOCV)
DRT miss rate	On-track driving	$r = .71$
VOT Total Shutter Open Time	Stationary vehicle	$r = .65$
Tracking task error	Stationary (simulated driving)	$r = .81$

4. Discussion

The central question this paper addresses is whether a count of discrete driver inputs captures enough information about an IVIS task to predict how cognitively and visually demanding that task is. The present study, across three independent criterion measures and two assessment contexts, confirms this to be the case. Step count predicted on-track cognitive load, stationary visual demand, and simulated-driving tracking error with cross-validated correlations ranging from r = .65 to .81.

The cross-context stability finding (r > .67) is a prerequisite for practical application of the measure. It shows step counts obtained in a parked vehicle are consistent with those obtained on a closed track, which in turn demonstrates that the criterion validity established against on-track DRT data transfers to a stationary testing setting. These findings support the use of step count as a valid distraction inference without the need for a proving ground nor a safety driver.

Model selection raised an important distinction between statistical optimality and utility. The more complex model offered a small cross-validated improvement for the DRT, but its participant-varying step coefficient would require prior observations from each new vehicle to generate predictions. The simpler random-intercept model, with a single fixed step-to-distraction slope, can be applied to any future IVIS system from the step count alone. This enables the metric to be applied at scale, by a non-expert observer, in a parked vehicle, in approximately 45–50 minutes, without any instrumentation, and its output can be interpreted against the cost functions derived here without further modelling.

The non-linear accumulation of visual cost across steps has direct implications for interface design. Given that later interactions in a sequence impose disproportionately greater eyes-off-road demand, the total attentional burden of a complex IVIS task is roughly exponential. Designs that reduce step depth through smarter defaults, predictive input recognition, and streamlined menu structures offer a compounding safety benefit. Future work should examine whether cost functions generalise to AI-driven voice interfaces, haptic surfaces, and augmented-reality displays, and whether periodic re-validation against on-track data is needed as IVIS technology continues to evolve.

5. Conclusions

Step count, the number of discrete driver inputs required to complete an IVIS task, is a valid and cross-validated predictor of cognitive load and visual attentional demand, confirmed against DRT, VOT, and tracking task criterion measures. Its cross-context stability means it can be obtained in a stationary vehicle and applied to novel IVIS systems without novel on-track data collection. The model-derived cost functions allow a step count to be directly translated into quantitative estimates of cognitive and visual attentional cost, providing the empirical foundation for a consumer-facing IVIS safety rating that requires no specialist expertise to administer.

6. Acknowledgments

This research was funded by the Australian Automobile Association (AAA) as part of the AAA Road Safety Research Program (Project ID: DD2021-02) and the research approach and findings do not necessarily reflect the views of the AAA or its member clubs. In accordance with the contractual arrangements for this research, the AAA has ownership of the developed IP and de-identified data from the research and any use requires the express written approval of the AAA. The authors acknowledge ABMARC for engineering and technical support, and the Australian Automotive Research Centre (AARC) for provision of the proving ground. The authors thank Professor Dave Strayer and Dr Joel Cooper (Red Scientific) for their contributions to the test protocol, study design, and analysis plan.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boyle, L. N., Lee, J. D., Guo, H., & Jenness, J. W. (2020). Tactile detection response task: Metrics for assessing drivers' cognitive workload. *Human Factors*, *62*(7), 1115–1128. https://doi.org/10.1177/0018720819872854

International Organization for Standardization. (2016). *Road vehicles: Transport information and control systems: Detection-response task (DRT) for assessing attentional effects of cognitive load in driving* (ISO Standard No. 17488:2016). https://www.iso.org/standard/59887.html

Getty, D., Biondi, F., Morgan, S. D., Cooper, J. M., & Strayer, D. L. (2018). The effects of voice system design components on driver workload. *Transportation Research Record*, *2672*(35), 10–19. https://doi.org/10.1177/0361198118777382

McDonnell, A. S., Imberger, K., Poulter, C., & Cooper, J. M. (2021). The power and sensitivity of four core driver workload measures for benchmarking the distraction potential of new driver vehicle interfaces. *Transportation Research Part F: Traffic Psychology and Behaviour*, *83*, 99–117. https://doi.org/10.1016/j.trf.2021.09.019

Metz, B., Landau, A., & Just, M. (2014). Frequency of secondary tasks in driving: Results from naturalistic driving data. *Safety Science*, *68*, 195–203. https://doi.org/10.1016/j.ssci.2014.03.009

Strayer, D. L., Cooper, J. M., & Castro, S. C. (2016). Extending the Detection Response Task to simultaneously measure cognitive and visual task demands. *Human Factors*, *58*(6), 869–875. https://doi.org/10.1177/0018720816642566

Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D., & Biondi, F. (2017). *Visual and cognitive demands of using in-vehicle infotainment systems*. AAA Foundation for Traffic Safety.

Strayer, D. L., Cooper, J. M., McCarty, M. M., Getty, D. J., Wheatley, C. L., Motzkus, C. J., Goethe, R. M., Biondi, F., & Horrey, W. J. (2019). Visual and cognitive demands of CarPlay, Android Auto, and five native infotainment systems. *Human Factors*, *61*(8), 1371–1386. https://doi.org/10.1177/0018720819836575

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Human*

Factors, *57*(8), 1300–1324. https://doi.org/10.1177/0018720815575149

Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). Cognitive distraction while multitasking in the automobile. In *Psychology of Learning and Motivation* (Vol. 54, pp. 29–58). Academic Press. https://doi.org/10.1016/B978-0-12-385527-5.00002-4

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Distracted driving – a systematic review with regard to resources used

Mark Vollrath[1], Maximilian Borowy[1],Sarah Helweg[1], Mette Møller[2*]

[1] Technische Universität Braunschweig, Department of Traffic and Engineering Psychology, Gaußstr. 23 38108 Braunschweig, Germany
[2] Technical University of Denmark (DTU), Division of Transportation Science, Akademivej 358, 2800 Kgs. Lyngby, Denmark
[*]mette@dtu.dk

Abstract: Engagement in secondary tasks while driving has surged due to mobile technology, yet the specific task characteristics that most detrimentally impact driver performance remain debated. This study provides a systematic overview of the negative effects of secondary task engagement on driving performance. Applying an adapted PRISMA methodology, 79 simulator-based studies (published 2015–2025) were analyzed. Data from 764 results were coded based on an extension of Wickens' Multiple Resource Theory, categorizing tasks by input modality (visual and acoustic), cognition required, and response type (manual and verbal). Effects on longitudinal control, lateral control, and reaction time were quantified using "vote counting" and logistic regression.
Complex visual input significantly increases negative effects (rising from 27% to 42%), especially when paired with complex cognitive processing (reaching 45%). Acoustic input, by contrast, causes fewer negative impacts. Combining complex visual input with manual reactions increases negative effects, particularly for tasks lacking visual control. Complex cognition alone does not impair driving unless it involves manual reactions, which compete for the same motor planning and execution resources used in driving (steering, braking). Contrary to expectations, complex verbal output (e.g., phone calls, voice messages) sharply increases negative effects (up to 48%), nearly doubling interference. Adding visual demands pushes negative effects close to 60%.
Tasks presenting purely acoustic input (e.g., listening to podcasts) yield the fewest negative effects (5%). Interactivity should be limited to simple acoustic signals and basic manual or voice commands. Complex visual-cognitive-manual tasks generate the highest interference by overloading shared processing resources. Further controlled laboratory research is recommended to isolate specific drivers of these negative effects.

1. Introduction

Previous studies have found distraction to be a leading driver-related factor in fatal crashes (Bakhit et al., 2018) especially for younger and older drivers (Guo et al., 2017). However, it remains unclear which characteristics of secondary tasks have the most detrimental effect on driver performance. Furthermore, previous results are mixed showing different effects in different situations and even reduced crash risk in non-demanding situations requiring simple behavioral reactions such as braking (for an overview see Vollrath et al., 2021).

Therefore, based on a systematic literature review, the purpose of this study was to provide a detailed look at task characteristics and resources used by the tasks which are responsible for their negative effects while driving.

2. Method

2.1 Selecting papers

An adapted version of the PRISMA methodology (Page et al., 2021) was applied consisting of four phases (see Fig. 1): Identification, Screening, Eligibility, and Inclusion.

Identification – a systematic search was done in the Scopus database using the following search string: "distraction" OR "inattention" OR "secondary task") AND ("driving" OR "vehicle operation" OR "automobile" OR "motorist") AND ("simulator" OR "driving simulator" OR "virtual reality" OR "simulation". It was decided to focus on driving simulator studies because only this experimental setting allows a comparison of different secondary task conditions with a control condition with driving, only, in comparable and replicable driving situations (including other traffic). 1165 publications from 2015 to the beginning of 2025 were included.

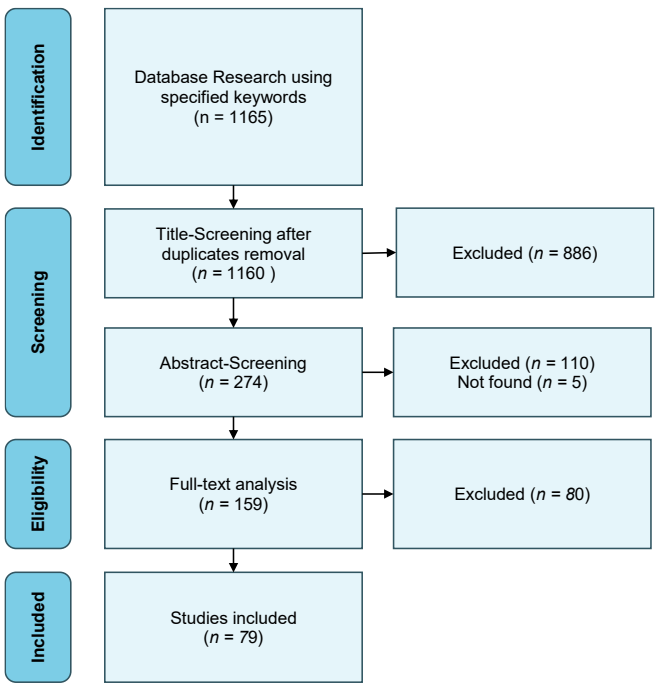


Fig. 1. *Overview of the selection process (PRISMA)*

Screening of abstracts to remove duplicates and publications who did *not* use an experimental set up, focused on car drivers (without automation), investigated task-based driver distrations and reported driving performance parameters. 886 studies were excluded based on the title and a further 115 studies based on the abstract.

Eligibility – a full-text analysis to exclude publications without comparison to a control condition, which focussed on environmental or infrastructure-related distractions, included a specialized sample or had methodological limitations preventing valid interpretation of the results.

Included - 79 publications who met all criteria were included.

2.2 Coding of results

The following information was coded: Study design (within subject, between subject), sample size, characteristics of driving scenario (e.g., urban, highway, speed limit), type of simulator, type of secondary task, type of distraction, driving performance parametres, and reported results. It was also coded if the results indicated improvement (e.g., shorter reaction time), deterioration (e.g., longer reaction time) or no effect on driving performance.

When examining the types of tasks more closely, we found that the resources required to do these tasks varied very strongly between the studies. For example, the task 'phoning' could include searching for a certain number and typing it, talking or just listening, or also talking about complex topics or simply confirming that one was listening. Thus, a task like 'phoning' could lead to completely different levels of distraction.

In order to examine that, a classification was developed extending Wickens' Multiple Resource Theory (Wickens, 2002). For each task, the demand with regard to 5 different resources required to perform this tasks was coded by two raters, resulting in 5 codings per task:

1) Visual input (none orlittle vs. complex)
2) Acoustic input (none or little vs. complex)
3) Cognition (none vs. little or complex)
4) Manual reactions (none vs. manual reactions without or with little visual control vs. manual reaction with more or complex visual control)
5) Verbal reactions (none or simple vs. complex verbal output or interaction)

For describing the effects on driving performance, the "vote counting" method (Bushman & Wang, 2009) was applied. For every parameter reported (e.g, average speed) it was noted whether there was a significant difference in this parameter when driving with a certain secondary task as compared to the baseline condition of just driving.

To summarize the effect of each secondary task, the percentage of significant negative findings was computed. To examine the effects of the demands in the different resources, a logistic regression was computed with the outcome 1 (significant negative effect) vs. 0 (no effect).

3. Results

3.1 Characteristics of the studies

Overall, 79 studies were included with 764 results. The study sample size ranged from 10 to 101 (M = 46). Most results (95%) came from a within-subjects design.

Of the 764 results, 34.4% used artificial tasks designed to use different ressources (e.g., visual perception) or different psychological functions (e.g., memory). The remaining results regarded realistic tasks (see Fig. 2).

Regarding the psychological resources, about half of the results concern tasks with complex visual input (52.2%). Nearly half of the tasks used complex acoustic input (44.2%). In 82.2% of the tasks used complex cognition was required. Regarding manual reactions, 39.1% used none, 20.3% used manual reactions with no or very little visual control and 40.6% required complex visual control. The majority (86.3%) used no or very simpel verbal reactions, but 13.7% required complex verbal output.

3.2 Effects of secondary tasks

Fig. 2 shows the negative effects per task. Writing or reading SMS and operating the smartphone while holding it in the hand have the highest number of significant negative effects (57%, 62% and 63%, respectively). The lowest number of effects are found when gaming was done (20%), when media were present (like listening to music, 23%). However, some of the results seem unlikely. Reading and writing SMS had 28% negative results while reading SMS alone had 62% and writing SMS 57%. This is due to the fact that the complexity of what to read and write was very different between the combined and single tasks.

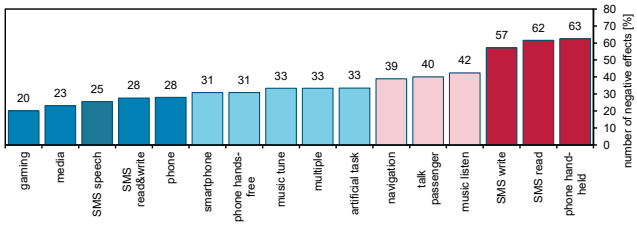


Fig. 2. *Percentage of negative effects per task.*

Thus, we examined the single and combined effects of using the different psychological resources in a logistic regression with finding a significant negative effect (1) or not (0) for each parameter examined as a criterion. First, all main effects were entered, then all 2-way interactions were added and all interactions were excluded which did not contribute (backwards procedure with WARD criterion). Table 1 gives the results of the final step of the regression.

Table 1. *Significant main effects and interactions for the logistic regression with finding significant negative effects as compared to baseline driving as the criterion.*

predictors	beta	WALD	p	OR
visual input	0.17	0.1	0.724	1.2
acoustic input	**-1.33**	**8.4**	**0.004**	**0.3**
cognition	0.36	1.8	0.182	1.4
manual reaction	**0.00**	**19.7**	**0.000**	**0.0**
verbal reaction	-0.02	0.0	0.947	1.0
visual input * cognition	**1.65**	**7.8**	**0.005**	**5.2**
visual input * manual reaction	**-1.93**	**9.1**	**0.003**	**0.1**
visual input * verbal reaction	**-1.97**	**5.1**	**0.024**	**0.1**
cognition * manual reaction	**1.73**	**5.4**	**0.020**	**5.6**

As main effects, acoustic input and manual reactions were significant. Visual input was significant in interaction with cognition, manual and verbal reactions. There was also an interaction between cognition and manual reactions. Fig.

3 shows the main effects. Complex visual input increases the percentage of negative effects from 27% without to 42% with complex visual input. In contrast, complex verbal input reduces the negative effects to 25% in comparison to tasks without verbal input (43%). There is also a strong confounding effect in that most tasks use either complex visual or complex acoustic input. Thus, most tasks without complex acoustic input use complex visual input instead.

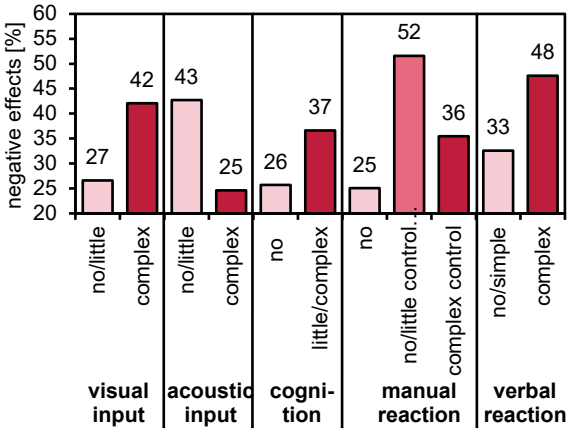


Fig. 3. *Percentage of negative effects for all resources examined.*

Complex cognition increases the negative effects to 37% as compared to 26% without. Manual reactions with no or little visual control lead to the largest number of negative effects (52%), but manual reactions with complex visual control (typing texts) also increase the percentage of negative effects to 36%. Finally, complex verbal reactions also increase the percentage of negative effects to 48%.

Fig. 4 shows the interaction of visual input and cognition. Basically, the combination of complex visual input and complex cognition leads to 45% negative effects and complex visual input without complex cognition lead to a smaller increase of only about 6% negative effects.

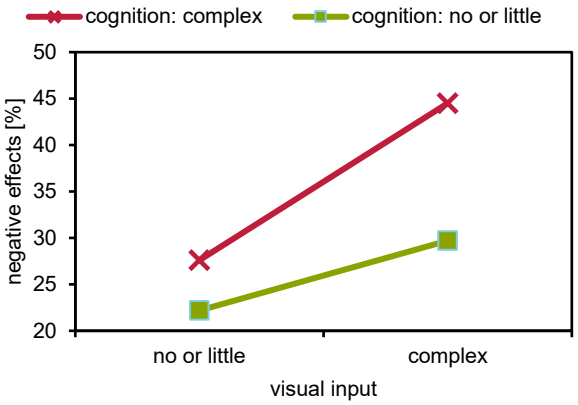


Fig. 4. *Interaction between visual input and cognition (percentage of negative effects).*

With regard to the interaction between visual input and manual reactions (see Fig. 5), complex visual input increases the number of negative effects in any combination, but especially in combination with manual reactions which do not need complex visual control.

Tasks with either no manual reactions or with manual reactions with complex visual control (typing on the phone) lead to the smallest percentage of negative reactions, only increasing when complex visual input is required. Thus, complex visual control itself does not seem to increase the number of negative effects.

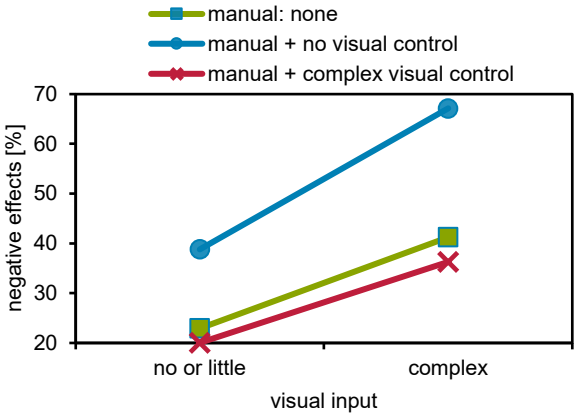


Fig. 5. *Interaction between visual input and manual reactions (percentage of negative effects).*

Finally, complex visual input and complex verbal output bot increases the number of negative effects and these two effects adds up, nearly reaching 60% in combination (see Fig. 6).

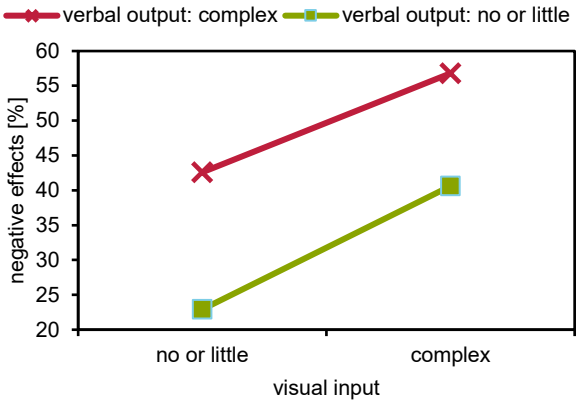


Fig. 6. *Interaction between visual input and verbal reactions (percentage of negative effects).*

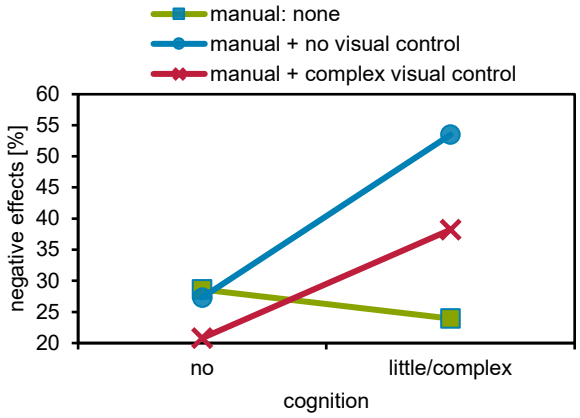


Fig. 7. *Interaction between cognition and manual reactions (percentage of negative effects).*

With regard to the interaction between cognition and manual reactions (see Fig. 7), complex cognition does not increase the number of negative effects alone when no manual reaction is required, but only when combined with manual reactions (with and without visual control). Again, this effect is stronger for manual reactions which do not require visual control. Thus, complex visual control alone does not lead to a larger number of negative effects.

4. Conclusions

As expected from Multiple Resource Theory, complex visual input leads to a strong increase of negative effects which is not seen when using complex acoustic input. However, this strong effect is mainly found in tasks combining complex visual input with complex cognition. Thus, it is not the visual input per se, but that it is also important to have to understand and process this input. On the other hand, when using no or little visual input, but acoustic input, it does not matter much whether there is a complex cognition required by the task. Thus, if a secondary task requires complex cognitive processing, the information for this task should be presented acoustically in order to avoid negative effects.

This increase of negative effects of secondary tasks with visual input is also found for all manual reactions. It is strongest when combined with manual reactions which do not require visual control. On the one hand, this effect of secondary task requiring manual reactions is also in accordance with the Multiple Resource Theory as driving also requires manual reactions. On the other hand, it was unexpected that tasks with manual reactions which require complex visual control (most of them writing on a virtual keyboard) do not differ from tasks requiring no manual output. One explanation could be that in these tasks examined the manual reactions were complex, but short, like typing only a few characters. However, the description of the tasks in the paper did not really allow to differentiate that.

When examining the effect of the combination of cognitive processing and manual reactions, there was a clear increase in negative effects for tasks which required manual reactions with and without visual control if the task also required complex cognition. If the tasks did not require manual reactions, there was no effect of tasks requiring complex cognition, but no manual reactions. Thus, complex cognition increases the number of negative effects if this leads to manual reactions, thus also requiring the planning of these reactions. This is in accordance with the Multiple Resource Theory as driving also requires planning and control of manual reactions (steering, accelerating, braking).

In contrary to the Multiple Resource Theory was the finding that complex verbal reactions increased the number of negative effects in the same manner as complex visual input. As driving does not require resources for verbal reactions, this should not be the case. Most of these tasks were phoning and sending speech messages via SMS. Even when no complex visual input was involved, the percentage of negative effects was about double as much as in tasks without complex verbal output. Adding visual input to the tasks with complex verbal output added up to nearly three times as much negative effects.

The tasks with the fewest number of negative effects (25%) used acoustic input, like listening to music or podcasts. It goes up to about 40% if the task requires complex cognitive processing or manual reactions. If complex verbal reactions are required, there are nearly 50% negative effects. With also manual reactions, 52% are reached. Thus, interactions with the driver should be best limited to acoustic signals to provide information or to require only simple manual or verbal reactions. Even using complex verbal reactions leads to an increase of negative effects. Thus, using voice-controlled tasks while driving these should stick to simple commands by the driver in order to minimize negative effects on driving.

Similar, tasks leading to the most negative effects involve complex visual input, a strong cognitive processing of this information and then manual reactions. These types of tasks use the same resources as driving and thus provide interference at all three levels of information processing.

However, these recommendations come from tasks which did not systematically vary and combine these different influencing factors. Moreover, it should be further examined what exactly leads to negative effects. Is it the amount of information or its complexity? How much visual control does the manual reaction really need? Further laboratory studies with a better control of task characteristics could be a way to answer these questions.

References[1]

Bahkit, P. R., Guo, B.B., & Ishak, S. (2018). Crash and near-crash risk assessment of distracted driving and engagement in secondary tasks: A naturalistic driving study. *Transportation Research Record,* 2672(38) 245–254. https://doi.dox.org/10.1177/0361198118772703

Bushman, B.J., & Wang, M.C. (2009). Vote-counting procedures in meta-analysis. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 207–220). Russell Sage Foundation.

Guo, F., Klauer, S.G., Fang, Y., Hankey, J. M., Antin, J.F., Perez, M.A., Lee, S.E., & Dingus, T.A. (2017). The effect of age on crash risk associated with driver distraction. *International Journal of Epidemiology*, 258-265. doi: 10.1093/ije/dyw234

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj, 372*.

Vollrath, M., Clifford, C., & Huemer, A. K. (2021). Even experienced phone users drive worse while texting – A driving simulator study. *Transportation Research Part F*, 78, 218-225. https://doi.org/10.1016/j.trf.2021.02.007

Wickens, C.D. (2002). Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3(2), 159-177.

[1] *The full reference list covering the literature included in the review is available upon request.*

Driver distraction at roadway event boundaries: a simulator investigation within an event segmentation framework

Avinoam Borowsky[1*], Noam Amit [1,2], Anat Meir [2]

[1] Ben-Gurion University of the Negev, Department of Industrial Engineering and Management, P.O.B. 653 Beer-Sheva 84105, Beer-Sheva, Israel
[2] Holon Institute of Technology, Faculty of Industrial Engineering and Technology Management, P.O.B 305 Holon 5810201, Holon, Israel
[*]borowsky@bgu.ac.il

Abstract: Driver distraction is a prominent factor contributing to road crashes, which motivates the development and implementation of driver monitoring systems (DMS) to identify driver distraction. However, existing DMSs often lack contextual sensitivity, leading to high false alarm rates and reduced trust. Drawing on the psychological Event Segmentation Theory (EST), this study investigated how drivers cope with a smartphone-related secondary task when navigating virtual road areas pre-defined as 'event boundaries' that are hypothesized to impose increased demands on central attentional resources. Twenty-one drivers navigated four virtual scenarios in a driving simulator while performing smartphone secondary tasks at varying time intervals relative to predefined road boundaries. A dual-lens camera recorded the participant's face and the external environment throughout the experiment. Findings provide preliminary evidence that drivers were more likely to brake suddenly and to postpone engaging with a secondary task when asked to do so closer to the boundary. Engagement lasted longer when tasks occurred farther from boundaries. These findings support the use of a context-aware framework in DMS design. The implications for integrating road context into driver assistance systems and their potential contribution to road safety are discussed.

Keywords: Driver distraction; Driver monitoring systems; Event Segmentation Theory; Roadway context; Secondary task engagement; Driving simulator

1. Introduction

Driver distraction is a leading contributor to traffic crashes, intensified by the proliferation of in-vehicle digital systems (Fasanmade, 2021; Michelaraki et al., 2023; Regan et al., 2008). Existing distraction-detection systems predominantly rely on driver-specific metrics while overlooking the broader driving context (Ayas et al., 2026; Sathyanarayana et al., 2011), which can lead to false positives, missed warnings, or delays, ultimately reducing system effectiveness (Mohammed, 2019). Qi (2020) emphasizes behavioural change through the 'three Es' (enactment, enforcement, and education) alongside technology adoption to address distraction. Driver Monitoring Systems (DMS) exemplify the latter by providing real-time alerts, which have been shown to mitigate distraction-related incidents (Scanlon et al., 2017; Masello et al., 2023). However, false alarms, which remain common byproducts of DMS, can undermine trust and discourage use, often with a greater impact than missed detections (Dixon et al., 2007; Tenhundfeld et al., 2020; Yang et al., 2017). Drivers who frequently engage in secondary tasks while driving rely on the driving context to regulate their behaviour across three hierarchical levels (Schömig & Metz, 2013; Zangi et al., 2022). At the strategic level, tasks are planned before the trip (e.g., calling a co-worker for updates). At the tactical level, drivers decide when to engage with the task based on road and traffic conditions (e.g., initiating the call when the road is straight and traffic is minimal). Finally, at the control level, once a driver begins engaging in a secondary task, they will allocate their attentional resources between the road and the task. Both decision and control levels require attentional resources for real-time decision-making, which may temporarily increase central attentional demands to the point where driving safety is compromised.

Quantifying the attentional demands imposed by the driving context at any given moment is difficult. A well-established memory-related theory, termed Event Segmentation Theory (EST), may provide a framework for identifying specific areas along the road where momentary increases in central attentional demands may occur. For example, driving for several minutes on a straight road, then approaching an intersection. The boundary may be perceived as the transition from a straight road to an intersection. Identifying these areas, which may require high levels of central attentional resources, in advance can provide DMS with input to prioritize the driving task in these areas by alerting the driver to focus on the road and avoid engaging in secondary tasks. Such an approach may help reduce false alarms. EST posits that one's continuous experience may be segmented into discrete events in long-term memory, with boundaries that demand central attentional resources and reduce reliance on automatic processes (Kurby & Zacks, 2008; Zacks et al., 2007; Ongchoco et al., 2023). In driving, transitions between different road settings (e.g., from a straight road into an intersection, approaching a sharp curve) may serve as event boundaries that disrupt continuity, require attentional reorientation, and limit engagement in secondary tasks (e.g., using a mobile phone).

Guided by the EST framework, we investigated how pre-designed roadway characteristics influence drivers' engagement with smartphone-related secondary tasks on the tactical and control levels (Schömig & Metz, 2013). The broader aim was to inform the design of intelligent DMSs that integrate driver mental state with roadway context to deliver selective, context-sensitive alerts. In addition, we postulated that engagement with a smartphone closer to the boundary would result in greater interference with the driving task because attentional resources would be divided between the secondary task and the driving task, where substantial

attentional resources are needed closer to the boundary to update the new driving-setting model. Notably, while most countries prohibit using a smartphone while driving, empirical data confirm that drivers often do so nevertheless (Huemer et al., 2018), with factors such as urgent work communications, social media notifications, and incoming calls ranking among the primary catalysts for this form of distracted driving (Calleo et al., 2025).

2. Method

2.1 Participants

Twenty-one students (10 male, 11 female) from Ben-Gurion University of the Negev participated (age range: 25-29 years; M=26.48, SD=1.12). All participants held a valid driver's license for at least five years. Participation was compensated with one bonus point in the course 'Introduction to Human Factors Engineering'.

Participants were randomly assigned to one of three groups (n=7 per group) based on the timing of the secondary-task engagement request relative to an event boundary detected by a time-to-collision (TTC) sensor. An event boundary was defined as a salient change in road or environmental conditions, such as approaching a sharp turn or intersection, or encountering a traffic sign indicating a merging road on the right. In the near-boundary group, secondary tasks were presented 5s before the event boundary; in the far-boundary group, they were presented 15s before the event boundary. Participants in the control group were not presented with secondary tasks.

2.2 Apparatus

2.2.1 Driving simulator and Questionnaires:

The study was conducted using an RTI high-fidelity driving simulator (Fig. 1). The simulator consisted of a Cadillac sedan equipped with its original accelerator, brake, and steering controls.



Fig. 1. *Simulator vehicle (left and middle) and interior camera positioning (right)*



Fig. 2. *Simulator view of the right-merging road scenario (left), and the corresponding event boundary warning sign (right)*

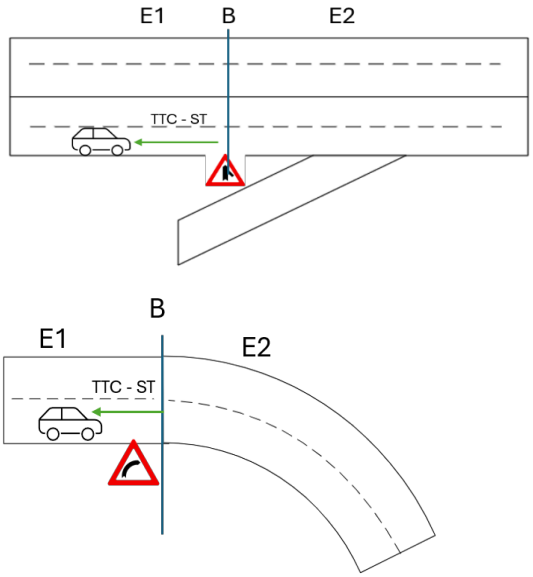


Fig. 3. *Illustration of the merging side road scenario (top) and sharp right curve (bottom). 'B' is the boundary, 'E1' and 'E2' are events 1 and 2. A time-to-collision (TTC) sensor positioned at the boundary triggered a request to engage in a secondary task (ST).*

The virtual driving environment was presented via three projectors on a curved screen with a diameter of 6.5m, providing a 165-degree field of view. Edge blending is performed using designated software (Wrapalizer, Inc.). The front visuals are complemented by synchronized audio feedback to enhance realism. In addition, two 10" LCDs are embedded in the side-view mirrors, and a rear-projector presents the rear view on the rear-view mirror. The steering system incorporates force feedback to emulate real-world steering resistance. Scenario design and implementation were conducted using SimVista software. A dual-lens dash camera by Provision was mounted on the vehicle's front windshield to unobtrusively record both the participant's face in the interior cabin and external roadway during the experiment, allowing subsequent observation of driver behaviour (e.g., gaze and head movements) and roadway conditions and external risk factors.

Participants completed a (1) demographic questionnaire and (2) the NASA-RTLX questionnaire to assess perceived workload and simulator experience.

2.2.2 Driving scenarios:

During the experiment, participants completed four driving scenarios (i.e., a sharp right curve, a right-merge road, a narrowing road, and a visually complex commercial environment). Scenarios were designed to examine the impact of secondary task engagement on drivers' performance as they drive through predefined road segment boundaries (e.g., transitioning from a rural environment with sparse traffic to an urban environment with dense traffic). Each scenario represented a distinct driving context with varying cognitive demands, enabling analysis of how environmental setting complexity interacts with concurrent task load to influence driver behaviour. The order of the scenarios was counterbalanced across participants

2.3 Procedure

Participants were randomly assigned to one of three conditions in a mixed design: near-boundary timing, far-boundary timing, or control. All participants completed four driving scenarios of varying complexity (see Figs. 2-3 for examples). In the experimental conditions, participants performed scenario-specific visual-manual secondary tasks on a mobile device: responding to a WhatsApp message with an emoji, dialling a single digit (1-9), or entering a countdown sequence from 9 to 1. Participants in the control condition completed the same driving scenarios without performing a secondary task. Each secondary task was assigned to a predefined scenario. For the WhatsApp task, the experimenter sent the message when the corresponding prompt appeared on the screen (Fig. 2). After completing each secondary task, participants were instructed to return to the device's home screen.

The tasks were presented on a smartphone mounted to an arm near the central stack. In the experimental groups, the task was presented either 15 seconds (far) or 5 seconds (near) before a predefined event boundary. Control participants drove without secondary tasks, but data was recorded at the same locations as in the other experimental conditions.

Each session lasted approximately 40 minutes. In the preparatory phase, participants signed an informed consent form, completed a demographic questionnaire, received standardized instructions on simulator operation and secondary tasks, and completed a brief road sign checkup to ensure familiarity while minimizing expectancy bias. Participants were instructed to drive as they normally would under similar real-world conditions and to complete secondary tasks promptly without compromising safety.

Participants first completed a brief training drive to familiarize themselves with the simulator and task requirements. The test phase comprised four driving scenarios (≈4 minutes each). Each scenario, except in the control condition, included a specific secondary task. At the conclusion, participants completed the NASA-RTLX mental workload questionnaire (Rubio et al., 2004).

3. Results

To understand the extent of the secondary task's interference with driving performance, we used the generalized linear mixed models (GLMM) framework, in which each regression model was fitted to the distribution of the dependent variable (e.g., logistic regression for binary variables, linear regression for normally distributed variables). The fixed effects in the initial models included the experimental condition (control, near boundary, far boundary), gender, driving experience (years), and their first-order interactions. In addition, participant ID was included as a random effect. The final model for each analysis was obtained via backward elimination. Post hoc pairwise comparisons were performed using Tukey-adjusted contrasts when necessary.

The dual-camera-based dependent variables (coded independently by two raters) aimed to capture various aspects of driving performance and task management. Lane departures indicated whether the vehicle veered out of its lane (coded as '0'=no or '1'=yes). Lane centring measured how well the vehicle was positioned within the lane (rated on a 5-point scale, with '1' indicating poor lane keeping and '5' representing optimal centring). Sudden braking indicated whether an abrupt braking event occurred (0=no, 1=yes). Task duration refers to the time taken to complete the secondary task, measured in seconds, with possible values starting from '0'. Finally, the task-postponement variable was a binary indicator of whether engagement with the secondary task began right when the text message appeared on the simulator screen or with a delay, coded as '0' for no and '1' for yes.

The main findings revealed that drivers initiated abrupt braking more often when the secondary task occurred near the boundary (estimated mean probability of a sudden brake = 0.542, SE = 0.105) than when no secondary task was present (0.0195, 0.002). No statistically significant difference was detected between the near- and far-boundary conditions. Furthermore, participants tended to engage with the secondary task for longer when it appeared farther from the boundary (estimated mean engagement duration=9.205s, 1.065s) than when it appeared closer to the boundary (7.4255s, 1.011s). In addition, participants tended to postpone their engagement with the secondary task more often when the task was near the boundary (estimated mean probability of postponing the task=0.161, 0.096) than when it was far from the boundary (0.0198, 0.002).

Lastly, although we do not report the analysis of vehicle dynamics in this abstract, Fig. 4 illustrates braking behaviour in the merging-road scenario shown in Fig. 3. Two interesting patterns were observed. First, participants who engaged with the secondary task closer to the boundary tended to delay their braking response (green line) beyond the point of conflict (merging road). Second, participants who engaged with the secondary task farther from the boundary tended to apply greater brake force when the merging-road sign became visible before the conflict area (i.e., the boundary).

4. Discussion

This study examined how engagement with a smartphone-related task in close proximity to a road area that places greater demands on central attentional processing affects driving performance. We adopted the EST psychological theory concept to design virtual road situations in a driving simulator that required a momentary increase in central attention demand.

The findings suggest that secondary-task engagement near an event boundary is associated with greater interference with driving performance, consistent with the proposition that such boundaries impose increased attentional demands. This momentary increase may be regulated by postponing engagement with the secondary task and prioritizing driving. Such regulation strategies are consistent with previous studies (Schomig & Metz, 2013; Zangi et al., 2022). There is also a large variability among drivers in how they regulate their engagement with the secondary task, which will be explored in future studies.

5. Conclusions

This study introduces a novel approach that uses the concept of EST to capture road areas where demands for central attention-processing resources may momentarily increase. Under such circumstances, the study showed that secondary tasks presented in close proximity to these areas (i.e., event boundaries) were associated with greater performance decrements than those presented farther away. Identifying such road areas can serve as context-sensitive features embedded in DMS to reduce false alarms, improve accuracy, and enhance safety. Incorporating context-aware information is a critical step toward advancing intelligent, adaptive driver support systems for robust distraction detection (Ayas et al., 2026).

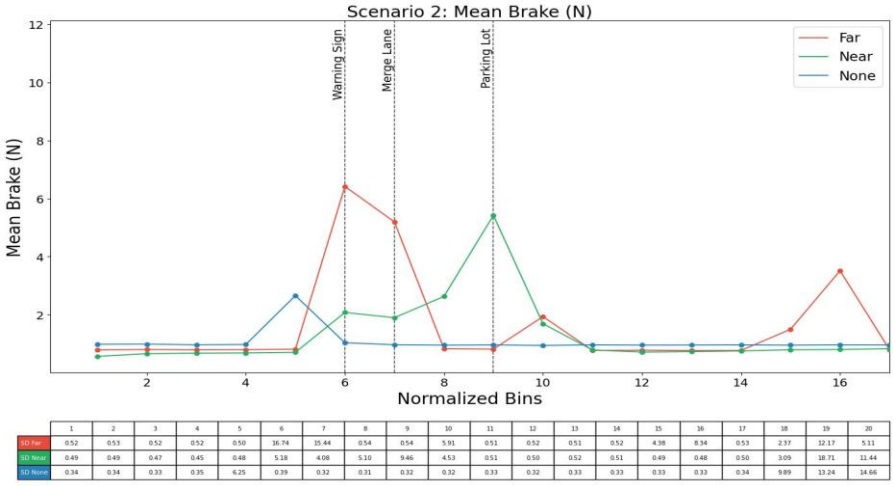


Fig. 4. *Illustration of the braking behaviour analysis in the right-merging road scenario*

6. Acknowledgments

This research was funded by the Israeli Ministry of Innovation, Science & Technology.

7. References

Ayas, S., Powell, M., & Donmez, B. (2026). Distraction mitigation through driver monitoring systems: a scoping review. Transportation Research Part F: Traffic Psychology and Behaviour, 121, 103701.

Calleo, Y., Di Zio, S., & Giuffrida, N. (2025). Assessing smartphone use while driving: A Real-Time Delphi study on risks and prevention strategies. Research in Transportation Business & Management, 62, 101469.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human factors*, 49(4), 564–572.

Fasanmade, A. (2021). A Context Aware Classification System for Monitoring Drivers' Distraction Levels (Doctoral dissertation, De Montfort University).

Huemer, A. K., Schumacher, M., Mennecke, M., & Vollrath, M. (2018). Systematic review of observational studies on secondary task engagement while driving. *Accident Analysis & Prevention, 119*, 225-236.

Kashevnik, A., Shchedrin, R., Kaiser, C., & Stocker, A. (2021). Driver distraction detection methods: A literature review and framework. *IEEE Access, 9,* 60063–60076.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in cognitive sciences, 12(2),* 72–79.

Masello, L., Sheehan, B., Castignani, G., Shannon, D., & Murphy, F. (2023). On the impact of advanced driver assistance systems on driving distraction and risky behaviour: An empirical analysis of Irish commercial drivers. *Accident Analysis & Prevention*, 183, 106969.

Michelaraki, E., Katrakazas, C., Kaiser, S., Brijs, T., & Yannis, G. (2023). Real-time monitoring of driver distraction: State-of-the-art and future insights. *Accident Analysis & Prevention*, 192, 107241.

Mohammed, M. M. (2019). Enhancing the Architecture of Context-Aware Driver Assistance Systems by Incorporating Insights from Naturalistic Driving Data.

Ongchoco, J. D. K., Walter-Terrill, R., & Scholl, B. J. (2023). Visual event boundaries restrict anchoring effects in decision-making—Proceedings of the National Academy of Sciences, 120(44), e2303883120.

Qi, Y., Vennu, R., & Pokhrel, R. (2020). Distracted driving: a literature review. FHWA-ICT-20-004.

Regan, M. A., Lee, J. D., & Young, K. (2008). Driver distraction: Theory, effects, and mitigation. CRC Press.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. Applied psychology, 53(1), 61–86.

Sathyanarayana, A., Boyraz, P., & Hansen, J. H. (2011). Information fusion for robust 'context and driver aware'active vehicle safety systems. *Information Fusion*, 12(4), 293–303.

Scanlon, J. M., Sherony, R., & Gabler, H. C. (2017). Injury mitigation estimates for an intersection driver assistance system in straight crossing path crashes in the United States. *Traffic injury prevention*, 18(sup1), S9-S17.

Schömig, N., & Metz, B. (2013). Three levels of situation awareness in driving with secondary tasks. Safety science, 56, 44–51.

Tenhundfeld, N. L., de Visser, E. J., Ries, A. J., Finomore, V. S., & Tossell, C. C. (2020). Trust and distrust of automated parking in a Tesla Model X. *Human factors*, 62(2), 194–210.

Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017, March). Evaluating effects of user experience and system transparency on trust in automation. *In Proceedings*

of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (pp. 408–416).

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological bulletin*, 133(2), 273.

Zhang, Z., Elahi, M. F., Domeyer, J., & Tian, R. (2025). Driver temporal segmentation of pedestrian crossing intentions during negotiations. *Transportation Research Part F: Traffic Psychology and Behaviour, 114,* 953–969.

Zangi, N., Srour-Zreik, R., Ridel, D., Chassidim, H., & Borowsky, A. (2022). Driver distraction and its effects on partially automated driving performance: A driving simulator study among young-experienced drivers. *Accident Analysis & Prevention*, 166, 106565.

Hands on the wheel, head in the clouds? Mind-Wandering Signatures in Simulated Manual Driving

Séverine Brelaud[1,2,3], Eleonora Mascotto[1], Numa Basilio[2], Tal Seidel Malkinson[3], Thomas Andrillon[1]

[1] Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France
[2] Renault Group, Human Factor Research Department, Technocenter, Guyancourt, France
[3] Université de Lorraine, CNRS, IMoPA, F-54000, Nancy, France

Abstract: Inattention, and mind-wandering (MW) in particular, poses a major risk for road safety, yet reliable behavioural markers during driving remain scarce. This study investigated behavioural and physiological signatures of spontaneous MW during a monotonous simulated manual driving task. Fifty-two healthy adults completed two 50-minute simulated sessions designed to elicit frequent MW, reporting their attentional state in response to auditory probes. EEG, ECG, eye-tracking, and driving behaviour data were continuously recorded and synchronised.
We focused on behavioural indices from the 10-second window preceding each probe (mean speed and speed variability, CV). Probe-locked analyses showed consistent effects of mental state: MW was linked to higher speed and markedly reduced speed variability across both urban and rural segments. By combining dense self-reports with synchronised measurements, we show that short timescale driving behaviour carries robust signatures of spontaneous MW. These behavioural markers offer a promising basis for real-time driver monitoring systems, with future multimodal analyses expected to further refine these signatures.

1. Introduction

Driver inattention, defined as insufficient or absent attention to activities essential for safe driving (Regan et al., 2011), is identified as the third leading contributing factor in road accidents, responsible for 14% of fatal crashes in France (ONISR, 2024). A significant component of driver inattention is mind-wandering (MW), a spontaneous cognitive state where attention shifts from the task at hand to self-generated thoughts and feelings (Smallwood & Schooler, 2015). MW occurs during approximately 50% of our waking life (Burdett et al., 2018) but increases with sleep deprivation (Poh et al., 2016). MW negatively impacts driving performance (Zhang & Kumada, 2018). It significantly increases the likelihood of being responsible for, or involved in, an accident (Galera et al., 2012), and is associated with risky driving behaviours (Lemercier et al., 2014; Yanko & Spalek, 2014). MW episodes are associated with detrimental changes in driving performance, including longer braking reaction times, increased speed, reduced following distance, and slower responses to peripheral events (Yanko & Spalek, 2014; Pepin et al., 2021). These effects are pronounced in monotonous or familiar driving conditions, which increase susceptibility to MW (Berthié et al., 2015; Burdett et al., 2019). Detecting such states during driving is therefore essential.

Although previous work has shown that off-task states can be detected through behavioural measures (e.g., lane keeping, speed, reaction times) and physiological markers (eye movements, EEG, heart rate, EDA) (He et al., 2011; Andrillon et al., 2021; Pepin et al., 2018; Mazeres et al., 2025), few studies have examined these modalities simultaneously. It is also unclear how MW during driving relate to more general changes in arousal and sleep pressure. A unified behavioural–physiological signature of MW remains to be established.

The present study examines the behavioural dynamics of spontaneous MW during simulated manual driving. We aim to identify reliable temporal markers of MW, expecting increased speed and reduced speed variability compared with task-focused state. By focusing on naturally occurring attentional fluctuations in a context where lapses carry immediate safety risks, this work extends previous findings and addresses an understudied driving scenario.

2. Method

2.1 Participants

Fifty-five healthy adults (mean age = 24.5 ± 4.3 years; 23 males) took part in the study. All were French speakers, with normal or corrected vision, no hearing impairments, and no reported sleep or vigilance disorders. Participants needed suitable hair for EEG and a valid driving licence. After data quality checks, three participants were excluded, yielding a final sample of 52. The study received ethical approval from the Comité d'Éthique de la Recherche de Sorbonne Université.

2.2 Driving scenario

A simplified automatic-transmission car simulator, consisting of a 42.5-inch screen, an office chair, a steering wheel, and two pedals was used. Participants completed a straight, monotonous driving task designed to induce frequent MW. Twelve events (red lights and pedestrian crossings) assessed attentional responsiveness. Urban and rural road sections alternated twice per session to avoid confounding road type with habituation, each with the same number of events. Rural (110 km/h) were longer than urban ones (70 km/h), spacing events further apart and increasing monotony.

2.3 Procedure

Participants were equipped with an electroencephalogram (EEG; 64-channel ActiChamp system with two AUX electrocardiogram (ECG) channels, ~250 Hz

during the task) and a PupilCore eye-tracker (~250 Hz). After calibration, they completed two 50-minute simulated driving sessions separated by a 20-minute eyes-closed rest to assess the effect of a brief recovery period on MW dynamics. EEG and ECG during the break were sampled at 500. Driving behaviour (speed, lane position, events) was sampled at 50 hz on a driving simulator. During driving, auditory probes occurred every 60–90 seconds (Yanko & Spalek, 2014), and participants used four buttons on the steering wheel to report (1) their current mental state (MW, task-focused, mind blanking (MB), external distraction), (2) thought immersion, and (3) alertness, yielding 76 reports per participant.

No responses were lost due to excessively fast reaction times (<150 ms). Five of the 110 completed sessions were removed due to recording issues.

All data streams were synchronised using Lab Streaming Layer (LSL).

2.4 Analysis

Mean and SEM were computed for each mental state across road types, with interindividual variability modelled using participant-level random effects. Non-parametric tests were used when the assumption of normality was violated. Driving metrics were extracted from the 10-second probe locked window preceding each auditory probe, providing speed (km/h) and speed variability (coefficient of variation, CV = SD/mean × 100%) (Yanko & Spalek, 2014; Baldwin et al., 2017).

3. Results

3.1 Analysis 1- Mental States x Road Type

We first verified that participants showed substantial variability in mental states during driving. Across the experiment, participants reported being task-focused 36.9 ± 2.5%, MW 34.1 ± 2.4%, MB 10.6 ± 1.6% and externally distracted 20.4 ± 1.6%. MW was slightly more frequent than task-focused during rural segments (37.5 ± 2.7% vs. 34.3 ± 2.6%), and less frequent on urban segments (30.2 ± 2.3% vs. 41.1 ± 2.4%) (Figure 1). We restricted analyses to task-focused (ON) vs. MW states, as the goal was to contrast focused driving with MW. Other states will be analysed in future work.

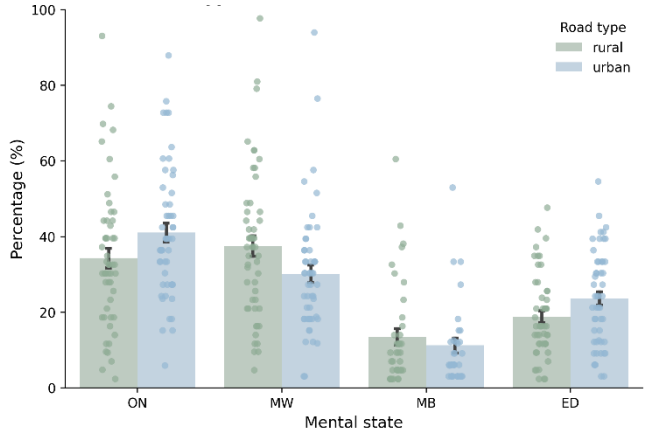


Figure 1: Distribution of mental states across road type. Dots represent probe-locked reports for each participant.

3.2 Analysis 2 – Speed: Mental State x Road Type

We fitted a linear mixed-effects model (LMM) with mental state (ON vs MW), road type (urban vs rural), and their interaction as fixed effects, and participant as a random intercept. Driving speed was higher on fast roads than on slow roads ($\beta = 19.44$, SE = 1.41, $z = 13.80$, $p < .001$). A smaller but significant main effect of mental state also emerged, with higher speeds during MW than during ON-task states ($\beta = 3.11$, SE = 1.42, $z = 2.19$, $p = .028$). These effects were qualified by a significant interaction ($\beta = 8.13$, SE = 1.99, $z = 4.08$, $p < .001$), indicating that the MW-related increase in speed was amplified on fast segments. Model-estimated values are shown in Figure 2.

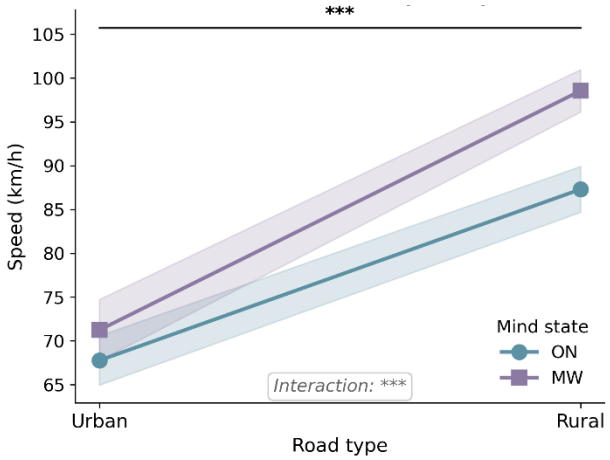


Figure 2: Model-estimated effects of mental state (ON vs MW) and road type (urban vs rural) on speed (km/h). Shaded areas show the 95% confidence interval (CI).

3.3 Analysis 3- Speed CV: Mental State x Road Type

To complement the participant-wise analyses, we fitted a linear mixed effects model (LMM) at the probe level with mental state (MW vs ON), road type (urban vs rural), and their interaction as fixed effects, and participant as a random intercept.

MW was associated with a significant reduction in CV across road types ($\beta = -4.78$, $p < .001$) (Figure 3), though the interaction between mind-state and road-type did not reach significance ($p = .14$), suggesting the MW effect on speed variability was rather consistent across road contexts.

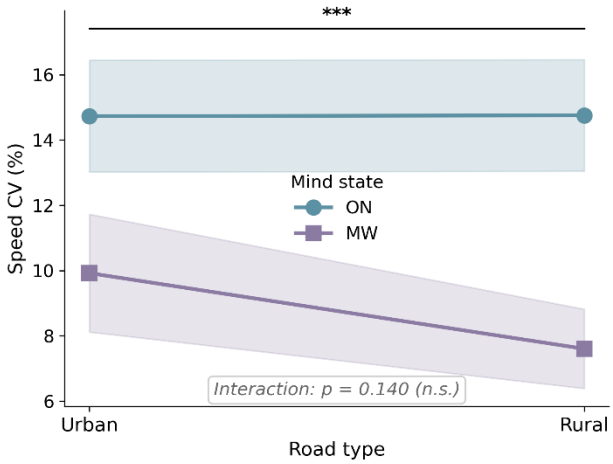


Figure 3: Model-estimated effects of mental state (ON vs MW) and road type (urban vs rural) on speed variability (CV). Shaded areas show the 95% CI.

4. Discussion

This study characterises spontaneous MW during simulated manual driving using probe locked behavioural measures. Consistent with previous reports that MW affects driving performance and responsiveness (Yanko & Spalek, 2014; Pepin et al., 2021; Lemercier et al., 2014), we observed clear behavioural signatures associated with internally oriented attention. At the probe level, MW was systematically linked to higher driving speed across both road contexts and to reduced speed variability (CV) in the seconds preceding MW reports.

Methodologically, the within participant, road type–specific aggregation isolates MW related fluctuations from baseline contextual differences. This is crucial as rural and urban segments differ speed and monotony, known to modulate MW prevalence (Berthié et al., 2015). Using CV ensured that variability differences did not simply reflect differing speed ranges across road types. While previous MW driving studies often emphasised ocular or physiological markers (He et al., 2011; Pepin et al., 2018; Andrillon et al., 2021), the present results show that short timescale behavioural dynamics tied to subjective experience are sufficient to recover a robust MW signal even when using a short time window (10s). The probe sampling frequency represents a good balance between sampling MW frequently enough to capture its occurrence and intensity (Seli et al., 2013), while still providing sufficient data for future detection-model development.

Importantly, the negative association between MW prevalence and speed variability echoes earlier findings that MW leads to reduced behavioural responsiveness and altered control strategies (Lemercier et al., 2014; Zhang & Kumada, 2018). By combining dense experience sampling with a multimodal acquisition framework, while analysing the behavioural stream alone, we demonstrate that variability-based indices already constitute sensitive and interpretable markers of attentional state. In future work, integrating physiological measures collected here will clarify whether these behavioural signatures are reinforced or temporally refined by multimodal evidence.

5. Conclusions

This study shows that spontaneous MW during simulated manual driving is reliably associated with higher mean speed and lower speed variability in the moments preceding self-reported MW episodes. These probe-locked behavioural signatures remain consistent across road contexts, indicating that short behavioural windows carry clear MW signatures, with potential for real-time monitoring of drivers' internal states. The multimodal recordings obtained in this experiment will enable future analyses to determine how physiological signals complement these behavioural signatures for improved MW detection.

6. Acknowledgments

The authors would like to express their gratitude to the participants of this study. This research was funded by a CIFRE PhD scholarship from Renault SAS to SB and a grant from the Fondation Vinci Autoroutes to TA.

References

Andrillon, T., Burns, A., Mackay, T., Windt, J., & Tsuchiya, N. (2021). Predicting lapses of attention with sleep-like slow waves. *Nature communications*, *12*(1), 3657. https://doi.org/10.1038/s41467-021-23890-7

Baldwin, C. L., Roberts, D. M., Barragan, D., Lee, J. D., Lerner, N., & Higgins, J. S. (2017). Detecting and quantifying mind wandering during simulated driving. *Frontiers in human neuroscience*, *11*, 406. https://doi.org/10.3389/fnhum.2017.00406

Berthié, G., Lemercier, C., Paubel, P. V., Cour, M., Fort, A., Galéra, C., ... & Maury, B. (2015). The restless mind while driving: drivers' thoughts behind the wheel. *Accident Analysis & Prevention*, *76*, 159-165. https://doi.org/10.1016/j.aap.2015.01.005

Burdett, B. R., Charlton, S. G., & Starkey, N. J. (2019). Mind wandering during everyday driving: An on-road study. *Accident Analysis & Prevention*, *122*, 76-84. https://doi.org/10.1016/j.aap.2018.10.001

Burdett, Bridget R. D., Samuel G. Charlton, et Nicola J. Starkey. 2018. « Inside the commuting driver's wandering mind ». Transportation Research Part F: Traffic Psychology and Behaviour, Special Issue on Everyday Driving, 57 (août):59-74. https://doi.org/10.1016/j.trf.2017.11.002.

Galera, C., L. Orriols, K. M'Bailara, M. Laborey, B. Contrand, R. Ribereau-Gayon, F. Masson, et al. 2012. « Mind Wandering and Driving: Responsibility Case-Control Study ». *BMJ* 345 (dec13 8): e8105-e8105. https://doi.org/10.1136/bmj.e8105.

He, J., Becic, E., Lee, Y. C., & McCarley, J. S. (2011). Mind wandering behind the wheel: Performance and oculomotor correlates. *Human factors*, *53*(1), 13-21. https://doi.org/10.1177/0018720810391530

Lemercier, Céline, Christelle Pêcher, Gaëlle Berthié, Benoit Valéry, Vanessa Vidal, Pierre-Vincent Paubel, Maurice Cour, et al. 2014. « Inattention behind the wheel: How factual internal thoughts impact attentional control while driving ». *Safety Science* 62 (février):279-85. https://doi.org/10.1016/j.ssci.2013.08.011.

Mazeres, F., Fort, A., Lavril, Y. *et al.* The impact of mind wandering on physiological measures in delegated driving. *Eur. Transp. Res. Rev.* 17, 42 (2025). https://doi.org/10.1186/s12544-025-00741-w

ONISR. (2024). *Accidentalité routière: Bilan définitif 2024*. https://www.onisr.securite-routiere.gouv.fr/sites/default/files/2025-05/2025%2005%2028_ONISR_Accidentalit%C3%A9_Bilan_d%C3%A9finitif_2024.pdf

Pepin, G., Fort, A., Jallais, C., Moreau, F., Ndiaye, D., Navarro, J., & Gabaude, C. (2021). Impact of mind-wandering on visual information processing while driving: An electrophysiological study. *Applied Cognitive Psychology*, *35*(2), 508-516. https://doi.org/10.1002/acp.3773

Pepin, G., Malin, S., Jallais, C., Moreau, F., Fort, A., Navarro, J., ... & Gabaude, C. (2018). Do distinct mind wandering differently disrupt drivers? Interpretation of

physiological and behavioral pattern with a data triangulation method. *Consciousness and cognition*, *62*, 69-81. https://doi.org/10.1016/j.concog.2018.04.006

Poh, J. H., Chong, P. L., & Chee, M. W. (2016). Sleepless night, restless mind: Effects of sleep deprivation on mind wandering. *Journal of experimental psychology: general*, *145*(10), 1312. https://doi.org/10.1037/xge0000207

Regan, M. A., Hallett, C., & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident analysis & prevention*, *43*(5), 1771-1781. https://doi.org/10.1016/j.aap.2011.04.008

Seli, P., Carriere, J. S., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in psychology*, *4*, 430. https://doi.org/10.3389/fpsyg.2013.00430

Smallwood, Jonathan, et Jonathan W. Schooler. 2015. « The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness ». *Annual Review of Psychology* 66 (1): 487-518. https://doi.org/10.1146/annurev-psych-010814-015331.

Yanko, Matthew R, et Thomas M Spalek. 2014. « Driving With the Wandering Mind: The Effect That Mind-Wandering Has on Driving Performance ». *Human Factors*. https://doi.org/10.1177/0018720813495280

Zhang, Y., & Kumada, T. (2018). Automatic detection of mind wandering in a simulated driving task with behavioral measures. *PloS one*, *13*(11), e0207092. https://doi.org/10.1371/journal.pone.0207092

The impact of non-driving related task engagement on the attentional processing of a novel unexpected auditory information

Yannick Lavril [1,*], Christophe Jallais [1], Clément Bougard [2], Alexandra Fort[1]

[1] Laboratory of Ergonomics and Cognitive Science for Transport (LESCOT), Université Gustave Eiffel, Cité des Mobilités, 25 av. François Mitterrand, 69500 Bron, France
[2] Stellantis – Creative center, Vélizy-Villacoublay, France
[*] yannick.lavril@univ-eiffel.fr

Abstract: Drivers are likely to engage in a non-driving related tasks (NDRTs) during delegated driving. However, even with SAE level 4 vehicles, drivers should ask or be asked by the system to regain control, requiring them to retain residual attention to process changes in the driving environment such as an auditory alert. This research aims to determine how three NDRTs and the level of engagement in them impact the attentional processing of unexpected, task-irrelevant novel sounds using event-related potentials (ERPs). Results demonstrate that ERP component amplitudes differed between the three NDRTs, especially one of them reducing both involuntary attention capture and the reorienting to the ongoing NDRT, whereas the engagement level manipulation had no effect. This suggests that the nature of NDRT may affect residual attention to unexpected auditory information.

1. Introduction

Depending on the level of vehicle automation, drivers may delegate driving to the system, and be willing to engage in non-driving related tasks (NDRTs) such as reading, using a smartphone, watching videos or observing the environment (Wilson et al., 2022). These NDRTs could become the main task during delegated driving. However, until SAE Level 4-5 vehicles are deployed, drivers will need to process unexpected auditory stimuli such as a takeover request. Responding adequately to such a stimulus requires that attention be first captured involuntarily, then reoriented away from the NDRT. Yet, how attention is involuntarily captured and reoriented by unexpected auditory stimuli when engaging in an NDRT remains understudied.

Assessing these attentional displacements during a task requires a method, such as a passive auditory oddball (Papanicolaou & Johnstone, 1984), that probes attention without adding any secondary task. This paradigm consists of ignoring task-irrelevant stimuli such as frequent standard tones and novel sounds, while brain activity is recorded. Electroencephalography (EEG) and the associated event-related potentials (ERPs) can be used to estimate the automatic processing of these sounds and the attentional dynamics. Three ERP components were mainly studied because they reflect different stages of stimulus processing when a novel stimulus disrupts a sequence of standard tones: (1) the P3a, linked to the involuntary capture of attention due to the deviance from the standard stimulus; (2) the novelty P3 reflecting the additional resources to processing novelty; (3) the RON, reflecting the disengagement from the stimulus and the attentional return to the primary task.

Several studies have shown that P3a, novelty P3 and RON amplitudes could decrease with increasing complexity during tasks (Ke et al., 2021; Scheer et al., 2016). These amplitude variations could be due to an increase in cognitive workload, and therefore indicate fewer residual attentional resources to process irrelevant stimuli. However, such effects are not always consistent across studies. Increasing task complexity could be accompanied by an increase in cognitive workload, but only if the individual mobilizes their resources and fully engages in the task. This engagement is based on the individual's motivation to produce an effortful investment to achieve a goal (Dehais et al., 2020). If a task is either unchallenging or too challenging, individuals may disengage and allocate fewer resources to the current task (Fairclough et al., 2013). This concept of task engagement seems to be central to understanding how drivers allocate their attentional resources during an NDRT.

While attentional resources have been compared across automation levels (Figalová et al., 2024), to our knowledge no study has investigated how engagement in different naturalistic NDRTs interacts with attentional dynamics. This study aims to determine how the level of engagement in three NDRTs impacts the processing of unexpected, task-irrelevant novel sounds and the attentional displacement using ERPs. As a first step, it was conducted without a driving simulator, to isolate the effects of NDRTs and engagement on attentional dynamics. Due to greater resource mobilisation, we expect to observe fewer residual attentional resources when the NDRT performed under conditions of high engagement as reflected by lower P3a, novelty P3 and RON amplitudes.

2. Method

2.1 Participants

Thirty-nine French right-handed volunteers participated in this study (21 women, age=26.41, SD=5.04) All participants reported having normal or corrected-to-normal vision, normal hearing, at least three years of driving experience, and no attentional, neurological diseases.

2.2 Non-Driving Related Tasks

Three different NDRTs were used: 'Video' with French subtitles; 'Reading', refering to history-related texts comprising at least 3,000 words; 'Observation', where participants who were not required to drive, had to look at

custom-made highway driving videos. Each NDRT lasted 10 minutes and none of them included a soundtrack.

2.3 Auditory stimuli

During each condition, participants were probed with 400 task-irrelevant auditory stimuli, including 340 repetitions of a standard tone (350 Hz) and 60 novel environmental sounds (e.g., thunder, etc.) selected from the New York State Psychiatric Institute database (Fabiani et al., 1996). The inter-stimulus interval was fixed at 1000 ms. The auditory stimuli lasted on average 331 ms (SD= 64.6 ms) and were presented at 60 dB SPL.

2.4 Engagement and Evaluation Context

Engagement was manipulated through the evaluation context, as anticipating a test elicits a greater engagement. Before each NDRT, participants were told that they would be questioned on its content at the end (Evaluated) or not (Non-evaluated).

2.5 Procedure

Participants were seated in a chair in front of a table equipped with a touch screen and speakers, positioned in front of a larger display screen. They completed the six experimental conditions (3 NDRTs x 2 Evaluation contexts) in a counterbalanced order. They were instructed not to pay attention to any auditory stimuli. Then, all participants answered the subjective questionnaires (motivation, NASA-TLX).



Fig 1: *Participant performing the driving environment observation task.*

2.6 Measures and Analysis

2.6.1 EEG:

EEG was recorded using the LiveAmp system (32 electrodes) at 500 Hz. EEG signal processing and ERP extraction were conducted using EEGLAB. To disentangle highly overlapping component a Principal Component Analysis was performed through ERP PCA toolkit. We conducted within-subjects Robust ANOVA (3 NDRTs x 2 Evaluation Context) on the difference novel minus standard amplitude for P3a, novelty P3 and RON components.

2.6.2 Subjective measures:

Motivation was rated on 7-points Likert scale. The 6 dimensions of the Raw NASA-TLX were computed to evaluate cognitive workload. Repeated measured ANOVAs (3 x 2) were conducted.

Table 1 *Robust ANOVA results*

ERP Component	NDRT		Evaluation Context (EC)		NDRT X EC	
	p	***D***	***p***	***dR***	***p***	***D***
P3a	**.023**	.58	.79	.13	.93	.67
novelty P3	.068	.62	.61	.17	.50	.50
RON	**.001**	.68	.22	-.21	.55	.76

Effect sizes were reported with D (project distance D) for multi-level within group and dR (robust analog of Cohen's d) for two-level within group.

3. Results

3.1 ERPs

The results can be found in Table 1. Post-hoc comparisons (Bonferonni corrected) revealed a greater P3a amplitude in Observation (M=5.36, SD=2.69) than in Video (M=4.23, SD=2.43, p=.028, dR=.42). As shown in Figure 2B, RON amplitude was smaller in Video (M=-.13, SD=1.51) than in Reading (M=-.97, SD=1.47, p=.004, dR=.54) and Observation (M=-.66, SD=1.82, p=.002, dR=.48).

3.2 Motivation

Analysis of motivation only revealed a significant effect of NDRT (p=.002, η^2p= .154). Post-hoc comparisons (Bonferroni corrected) revealed higher motivation in Video (M=5.769, SD=1.1229) than in Observation (M=4.949 SD=1.332, p<.001, d =.546).

3.3 NASA-TLX

The Evaluation Context indicated a main effect on mental (p<.001, η^2p = .426), effort (p<.001, η^2p = .345), physical (p=.013, η^2p = .151), and temporal demands (p=.047, η^2p = .10), all higher in the presence of a related-NDRT questionnaire, while achievement was rated lower (p=.006, η^2p=.181).

The analysis revealed a main effect of NDRT on temporal (p<.001, η^2p=.435) and mental demand (p=.037, η^2p=.091). Post-hoc comparisons revealed that temporal demand was higher in Video (M=9.885, SD=4.595) than in Observation (M=4.846, SD=3.453, p<.001, d=1.115) and Reading (M=5.487, SD=3.783, p<.001, d=1.115). Mental demand was only higher in Reading (M=12.526, SD=3.665) than in Observation (M=10.346, SD=4.401, p=0.022, d=0.468).

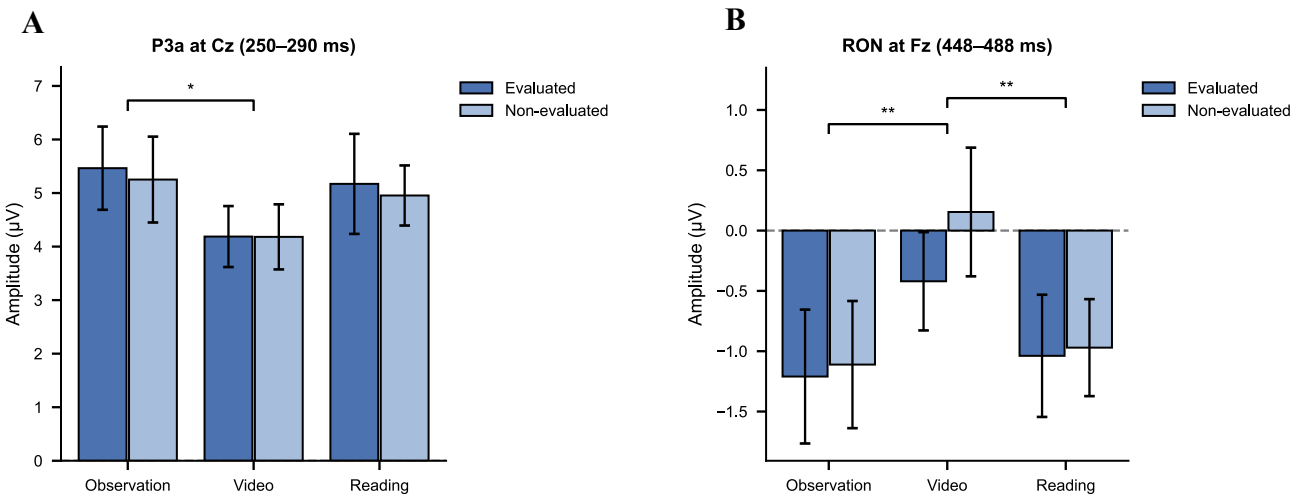


Fig 2: *Novel-minus-standard amplitude differences of the PCA factor scores for the P3a (Panel A) and RON (Panel B) components in each experimental condition. Error bars represent 95% within-subject confidence intervals. Brackets indicate significant comparisons (* p < .05, ** p < .01).*

4. Discussion

This study aimed to investigate how the level of engagement and the type of NDRT impact the processing of novel unexpected sounds and attentional displacement.

Although the announced questionnaire on the NDRT content increased perceived workload, it did not affect task engagement, as self-reported motivation remained unchanged. Nor did it impact attention displacement to unexpected novel sounds, as ERP components were unaffected.

Nonetheless, the type of NDRT modulated the nature of cognitive workload. Indeed, reading text imposed a higher mental demand than the observation task. The video task involved a higher temporal demand than reading text or observing, probably due to the uninterruptible nature of the task. Interestingly, P3a amplitude was lower in the Video than in Observation, indicating reduced involuntary attention capture by deviance and thus lower residual attention. Novelty P3, however, was not modulated by the type of NDRT, suggesting preserved novelty processing. RON amplitude was lower in Video than in the other two conditions, indicating reduced reorienting attention to the ongoing task. This may partly result from the reduced capture in Video, which required fewer resources to reorient attention to the video. It may also reflect the uninterruptible nature of the task, which prevents attention from fully disengaging.

5. Conclusions

These results contribute to a better understanding of how different NDRTs affect the processing of novel unexpected, task-irrelevant auditory information and attention dynamics. They also demonstrate the valuable insight that ERPs provide into residual attention during an NDRT. Further research will be necessary to determine whether these effects persist in a delegated driving simulator, whether they extend to driving-relevant sounds such as takeover request and whether they have behavioural consequences on takeover.

6. Acknowledgments

This research was conducted with the financial support of Stellantis.

References

Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00268

Fabiani, M., Kazmerski, V. A., Cycowicz, Y. M., & Friedman, D. (1996). Naming norms for brief environmental sounds : Effects of age and dementia. *Psychophysiology*, *33*(4), 462-475. https://doi.org/10.1111/j.1469-8986.1996.tb01072.x

Fairclough, S. H., Gilleade, K., Ewing, K. C., & Roberts, J. (2013). Capturing user engagement via psychophysiology : Measures and mechanisms for biocybernetic adaptation. *International Journal of Autonomous and Adaptive Communications Systems*, *6*(1), 63. https://doi.org/10.1504/IJAACS.2013.050694

Figalová, N., Bieg, H.-J., Reiser, J. E., Liu, Y.-C., Baumann, M., Chuang, L., & Pollatos, O. (2024). From Driver to Supervisor : Comparing Cognitive Load and EEG-Based Attentional Resource Allocation Across Automation Levels. *International Journal of Human-Computer Studies*, *182*, 103169. https://doi.org/10.1016/j.ijhcs.2023.103169

Ke, Y., Jiang, T., Liu, S., Cao, Y., Jiao, X., Jiang, J., & Ming, D. (2021). Cross-Task Consistency of Electroencephalography-Based Mental Workload Indicators : Comparisons Between Power Spectral Density and Task-Irrelevant Auditory Event-Related Potentials. *Frontiers in Neuroscience*, *15*, 703139. https://doi.org/10.3389/fnins.2021.703139

Papanicolaou, A. C., & Johnstone, J. (1984). Probe Evoked Potentials : Theory, Method and Applications. *International Journal of Neuroscience*, *24*(2),

107-131. https://doi.org/10.3109/00207458409089800

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2016). Steering Demands Diminish the Early-P3, Late-P3 and RON Components of the Event-Related Potential of Task-Irrelevant Environmental Sounds. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00073

Wilson, C., Gyi, D., Morris, A., Bateman, R., & Tanaka, H. (2022). Non-Driving Related tasks and journey types for future autonomous vehicle owners. *Transportation Research Part F: Traffic Psychology and Behaviour*, *85*, 150-160. https://doi.org/10.1016/j.trf.2022.01.004

Beyond Fixation: A Systematic Review of Peripheral Vision Use in Driving

Victor Meulenkamp[1*], Katja Kircher[2], & Christian Vater[1]

[1] Institute of Sport Science, University of Bern, Bremgartenstrasse 145, Bern, Switzerland
[2] The Swedish National Road and Transport Research Institute (VTI), S-581 95 Linkoping, Sweden
[*]victor.meulenkamp@unibe.ch

Abstract: Driving is a dynamic and sometimes complex task in which safety depends on the ability to perceive important information under time pressure. Given these constraints, peripheral vision plays a critical role in driving as it is not feasible to direct the fovea and fixate all relevant information simultaneously. This systematic review examines how perceptual, cognitive, and motor factors influence the use of peripheral vision in driving. Following PRISMA guidelines, a systematic search was conducted across PubMed, Scopus, and Web of Science. A total of 6,890 records were identified. After duplicate removal and abstract screening, 757 articles were eligible for full-text review. Studies were included if they investigated peripheral vision in dynamic driving settings and reported objective gaze measures. Eligible studies were categorized according to study design, manipulated factors, and driving task. This review will provide a structured framework for understanding peripheral vision in driving and provide a link to sport science, where athletes have to make decisions in similarly complex scenarios, also under time pressure.

1. Introduction

A driver approaches a busy intersection while maintaining lane position, monitoring a cyclist approaching from the right, checking the rear-view for overtaking vehicles, and anticipating a potential change of the traffic light, without continuously fixating each of these elements. This everyday situation illustrates the research topic of the current systematic review: the role of peripheral vision in driving. Driving is a dynamic activity in which drivers' actions not only depend on what they fixate, but also on what they perceive and monitor outside the fovea. Understanding how peripheral vision is used in driving is therefore critical for advancing research on driver distraction and inattention.

Peripheral vision refers to the visual field beyond 4 to 5 degrees around the fovea and differs from foveal vision in anatomical representation and perceptual capabilities (Larson & Loschky, 2009). Although the fovea comprises only about 1% of the retinal surface, it occupies 50% of the visual cortex (Curcio et al., 1990; Tootell et al., 1982), resulting in higher spatial resolution and color sensitivity compared to the periphery (Anstis, 1974; Strasburger et al., 2011; Abramov & Gordon, 1977; Gordon & Abramov, 1977; Hansen et al., 2009). With increasing eccentricity, visual acuity declines and object recognition becomes constrained by crowding, the phenomenon whereby objects become difficult to identify when surrounded by nearby stimuli (Bouma, 1970; Rosenholtz, 2016, 2020). These limitations can in principle be overcome by shifting gaze. However, eye movements (saccades) themselves have temporal costs and visual suppression, requiring to continuously trade off between relying on peripheral input and initiating a saccade to obtain high-resolution foveal information.

Peripheral vision is closely intertwined with attentional mechanisms, which determine how efficiently information can be extracted outside the fovea. Covert attention (i.e., attending to an object away from the point of gaze) enhances contrast sensitivity (Cameron et al., 2002; Carrasco, 2011), processing speed (Carrasco et al., 2006) and change detection in peripheral locations (Vater, 2019). The concept of the useful field of view (UFOV) captures the region around fixation from which information can be extracted without eye movements (Ringer et al., 2016). Importantly, UFOV is not fixed but can vary due to cognitive load (Gaspar et al., 2016), age (Underwood et al., 2005) and emotional state (Zhang et al., 2016). Under high attentional demands, peripheral processing efficiency can decrease, potentially narrowing the effective visual field.

Basic vision science shows that although peripheral vision is limited, it still provides useful information for many tasks. For example, it supports saccade planning, allows tracking of multiple objects at once, perceive the gist of a scene and perform object-recognition tasks(Wolfe et al., 2017). These findings suggest that peripheral vision is important for many everyday actions. However, its role is often underestimated in laboratory experiments, which typically use simplified stimuli and single-task settings that do not reflect real-world complexity. To fully understand how peripheral vision supports behavior, it is therefore necessary to examine its role in more complex and dynamic environments.

Applied research provides this broader perspective by examining peripheral vision in real-world tasks. According to the systematic review by Vater et al. (2022), peripheral vision supports four key functions across daily life tasks such as driving. These functionalities are monitoring the environment, presaccadic preview, managing saccade and eccentricity costs, and supporting action planning. Monitoring refers to the continuous extraction of information from the surroundings without direct fixation, such as detecting lane boundaries. Presaccadic preview allows observers to gather preliminary information about a peripheral object before fixating it, thereby facilitating efficient saccades. Finally, peripheral vision provides information that supports action planning, such as maintaining lane position through optic flow cues.

Although perceptual, cognitive and, motor constraints are well established in basic science, their combined

influence on peripheral vision use in dynamic driving remains unclear. Moreover, many driving studies rely on eye movement measures and assume that fixation reflects information processing (Harbluk et al., 2007; Strayer & Johnston, 2001), potentially underestimating the contribution of peripheral vision (Rosenholtz, 2016). Additionaly, unlike in sports research, where functional gaze strategies (e.g., gaze anchor, visual pivot, foveal spot) have been systematically described, no comparable framework exists to characterize patterns of peripheral vision usage in traffic contexts (Vater et al., 2020).

The present systematic review addresses these gaps by identifying and categorizing factors that influence peripheral vision use in driving. Specifically, we examine how perceptual, cognitive, and motor factors influence peripheral vision use in driving and how this affects gaze behavior and driver performance. By integrating findings across experimental paradigms and driving contexts, this review aims to provide a systematic framework that can be used in future research on driver distraction and inattention.

2. Method

2.1 Identification

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) procedure (Moher et al., 2009). The database search was conducted in April 2025 using Pubmed, Scopus and Web of Science. Only peer-reviewed articles written in English with accessible full texts were considered. Search terms were defined a priori and applied to titles, abstracts, and where available, keywords. A detailed description of the search strategy is provided in Appendix A (Table 1).

2.2 Screening

Using the predefined search strategy, a total of 6,890 records were identified from the selected databases. After removal of 2,089 duplicates, 4,801 records remained for title and abstract screening. The screening was conducted following a structured sequence based on the predefined criteria (see Appendix A, Table 1). First, records were screened for relevance to the target population, excluding studies involving non-human subjects (e.g., animals) and non-driver populations (e.g., cyclists, pedestrians, motorcyclists). Second, methodological criteria were applied, excluding studies conducted in inappropriate settings. Only studies involving dynamic driving-related visual content (i.e., real world driving, driving simulator, or screen-based/video-based driving contexts) were included, whereas static image tasks were excluded. Third, studies were assessed for the presence of objective gaze measures. Studies that did not include any form of gaze assessment were excluded. This process resulted in 757 records eligible for full-text screening.

2.3 Eligibility

Full texts were retrieved and screened according to more specific eligibility criteria focusing on the role of peripheral vision in driving. Studies were excluded if they focused solely on foveal vision without reference to peripheral vision. To ensure the relevance, the discussion section of each paper was screened for a list of terms, see Appendix A (Table 1). Studies that didn't address one of these terms were excluded. In addition, studies had to fit within the predefined categorization framework (study type, manipulation type, and driving task; see Appendix A, Table 2 & 3). Articles that could not be categorized across all three levels were excluded. So far, we excluded 526 full texts based on these criteria and will finish the screening by the end of May 2026. We expect to include 150 to 200 papers in this systematic review.

3. Preliminary results

The first results of the full-text screening indicate that driving requires close coordination of perceptual, cognitive, and motor processes. Drivers maintain lane position using peripheral optic flow and road-edge information (Kountouriotis et al., 2011), detect hazards that initially appear outside fixation (Shinoda et al., 2001), monitor mirrors while keeping gaze forward (Doshi & Trivedi, 2012), and respond to sudden events before fixating them (Huestegge & Böckler, 2016). Importantly, applied findings indicate that drivers sometimes initiate braking or steering responses prior to directly fixating a hazard, demonstrating that peripheral information can trigger action. Moreover, expertise and cognitive load influence how drivers balance peripheral monitoring and eye movements. Experienced drivers differ in their distribution of attention and gaze behavior compared to novices (Lamble et al., 1999; Summala et al., 1996; Underwood et al., 2003), and higher cognitive workload has been associated with reduced peripheral monitoring efficiency and functional visual narrowing, often described as visual tunnelling (Gaspar et al., 2016; Victor et al., 2005). These findings explain that peripheral vision plays a central role in safe driving performance. The full analysis of the included articles will provide a qualitative and quantitative summary of identified functionalities and discuss potential mechanisms linked to perceptual, cognitive and motor factors.

References

Abramov, I., & Gordon, J. (1977). Color vision in the peripheral retina. I. Spectral sensitivity. *Journal of the Optical Society of America*, *67*(2), 195–202. https://doi.org/10.1364/josa.67.000195

Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research*, *14*(7), 589–592. https://doi.org/10.1016/0042-6989(74)90049-2

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*(5241), 177–178. https://doi.org/10.1038/226177a0

Cameron, E. L., Tai, J. C., & Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. *Vision Research*, *42*(8), 949–967. https://doi.org/10.1016/s0042-6989(02)00039-1

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Carrasco, M., Giordano, A. M., & McElree, B. (2006). Attention speeds processing across eccentricity: Feature and conjunction searches. *Vision Research*, *46*(13), 2028–2040. https://doi.org/10.1016/j.visres.2005.12.015

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The*

Journal of Comparative Neurology, *292*(4), 497–523. https://doi.org/10.1002/cne.902920402

Doshi, A., & Trivedi, M. M. (2012). Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of Vision*, *12*(2), 9. https://doi.org/10.1167/12.2.9

Gaspar, J. G., Ward, N., Neider, M. B., Crowell, J., Carbonari, R., Kaczmarski, H., Ringer, R. V, Johnson, A. P., Kramer, A. F., & Loschky, L. C. (2016). Measuring the Useful Field of View during Simulated Driving with Gaze-Contingent Displays. *Human Factors*, *58*(4), 630–641. https://doi.org/10.1177/0018720816642092

Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. Ii. Hue and saturation. *Journal of the Optical Society of America*, *67*(2), 202–207. https://doi.org/10.1364/josa.67.000202

Hansen, T., Pracejus, L., & Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. *Journal of Vision*, *9*(4), 26. https://doi.org/10.1167/9.4.26

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis and Prevention*, *39*(2), 372–379. https://doi.org/10.1016/j.aap.2006.08.013

Huestegge, L., & Böckler, A. (2016). Out of the corner of the driver's eye: Peripheral processing of hazards in static traffic scenes. *Journal of Vision*, *16*(2), 11. https://doi.org/10.1167/16.2.11

Kountouriotis, G. K., Floyd, R. C., Gardner, P. H., Merat, N., & Wilkie, R. M. (2011). The role of gaze and road edge information during high-speed locomotion. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(3), 687–702. https://doi.org/10.1037/a0026123

Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics*, *42*(6), 807–815. https://doi.org/10.1080/001401399185306

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10), 6. https://doi.org/10.1167/9.10.6

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P. C., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Ringer, R. V., Throneburg, Z., Johnson, A. P., Kramer, A. F., & Loschky, L. C. (2016). Impairing the useful field of view in natural scenes: Tunnel vision versus general interference. *Journal of Vision*, *16*(2), 7. https://doi.org/10.1167/16.2.7

Rosenholtz, R. (2016). Capabilities and Limitations of Peripheral Vision. *Annual Review of Vision Science*, *2*, 437–457. https://doi.org/10.1146/annurev-vision-082114-035733

Rosenholtz, R. (2020). Demystifying visual awareness: Peripheral encoding plus limited decision complexity resolve the paradox of rich visual experience and curious perceptual failures. *Attention, Perception & Psychophysics*, *82*(3), 901–925. https://doi.org/10.3758/S13414-019-01968-1

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research*, *41*(25–26), 3535–3545. https://doi.org/10.1016/S0042-6989(01)00199-7

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision, 11*(5), 13. https://doi.org/10.1167/11.5.13

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*(6), 462–466. https://doi.org/10.1111/1467-9280.00386

Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors*, *38*(3), 442–451. https://doi.org/10.1518/001872096778701944

Tootell, R. B. H., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science (New York, N.Y.)*, *218*(4575), 900–901. https://doi.org/10.1126/SCIENCE.7134981

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*(6), 629–646. https://doi.org/10.1080/0014013031000090116

Underwood, G., Phelps, N., Wright, C., Van Loon, E., & Galpin, A. (2005). Eye fixation scanpaths of younger and older drivers in a hazard perception task. *Ophthalmic and Physiological Optics*, *25*(4), 346–356. https://doi.org/10.1111/j.1475-1313.2005.00290.x

Vater, C. (2019). How selective attention affects the detection of motion changes with peripheral vision in MOT. *Heliyon*, *5*(8), e02282. https://doi.org/10.1016/J.HELIYON.2019.E02282

Vater, C., Williams, A. M., & Hossner, E. J. (2020). What do we see out of the corner of our eye? The role of visual pivots and gaze anchors in sport. *International Review of Sport and Exercise Psychology*, *13*(1), 81–103. https://doi.org/10.1080/1750984X.2019.1582082

Vater, C., Wolfe, B., & Rosenholtz, R. (2022). Peripheral vision in real-world tasks: A systematic review. *Psychonomic Bulletin & Review*, *29*(5), 1531–1557. https://doi.org/10.3758/s13423-022-02117-w

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle

task difficulty. *Transportation Research Part F: Traffic Psychology and Behavior*, *8*(2), 167–190. https://doi.org/10.1016/j.trf.2005.04.014

Wolfe, B., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the Useful Field: Considering peripheral vision in driving. *Applied Ergonomics*, *65*, 316–325. https://doi.org/10.1016/j.apergo.2017.07.009

Zhang, T., Chan, A. H. S., Ba, Y., & Zhang, W. (2016). Situational driving anger, driving performance and allocation of visual attention. *Transportation Research Part F: Traffic Psychology and Behaviour*, *42*(2), 376–388. https://doi.org/10.1016/j.trf.2015.05.008

Appendix A

Table 1. *Search strategy used for including papers in the reviewed set. Criteria for 1) identification, 2) abstract screening and 3) eligibility evaluation.*

1) Identification (Method: Search databases)

Databases searched	*Inclusion criteria*	*Exclusion criteria*	*Search term used*
Pubmed (title, abstract)	Peer-reviewed	Conference abstracts	(peripheral* OR periphery OR attention OR useful field of view)
Scopus (title, abstract, keywords)	Full-text	Dissertations	AND (eye movement* OR eye tracking OR eye fixation OR gaze* OR visual search OR eccentricity OR visual signals OR distraction)
Web of Science (title, abstract, keywords)	English language	Book chapters	AND (driving OR driver*)
		Reviews	NOT (neural OR brain OR sleep OR cortex OR asd OR adhd)

2) Abstract Screening (Method: Exclude papers based on topic criteria)

Exclusion criteria	*Examples*
Animals	Cat, flies, apes
Vulnerable road users	Cyclists, pedestrians, motorcyclist
Clinical	Alzheimer, glaucoma
Drugs	Alcohol, cannabis
Fatigue	Drowsiness and fatigue detection
Ageing	Age-related attention impairments
Sport	Soccer, tennis
Wrong context	AI-driven, data-driven, experience-driven

3) Eligibility evaluation (Method: Inspect discussion section for relevant discussion terms and identify relevant study types)

Include papers that discuss one of these terms	*Include papers that belong to one of the relevant study and manipulation types (see Table 2)*
Peripher*	
Eccentric*	
Visual field	
Extrafove*	
Parafovea*	
Perifove*	
Covert	
Off-axis	
Side vision	
Useful field	
UFOV	
Ambient vision	
Global vision	

Table 2. *Framework for categorizing included studies on peripheral vision in driving.*

Level	Category	Definition (examples)
Study type	Experimental	Studies that manipulate visual or attentional conditions (see Table 3).
	Observational / exploratory	Studies that observe natural gaze or attention behavior without manipulation.
PV factor	Perceptual	Perceptual factors refer to studies which the goal is to manipulate visual or oculomotor constraints that determine what information can be perceive d in peripheral vision. For example, investigating the influence of time pressure, eccentricity (crowding), and the costs associated with eye movements.
	Cognitive	Cognitive factors refer to studies in which the goal is to manipulate attentional allocation, prediction, workload, or learned strategies influencing peripheral vision use. For example, exploring the impact of predictability, and cognitive dual-tasking.
	Motor	Motor factors refer to studies that examine the role of motor-related dual tasks.
Driving function	Steering / vehicle control	Driving tasks that involve continuous, ongoing control of the vehicle, requiring a correct regulation of vehicle position and motion over time (lane-keeping, passing cars, or speed-management).
	Detection / monitoring	Driving tasks in which the driver is required to notice, monitor or track objects or events in the traffic environment, often before an explicit response is required (hazards, road signs traffic signals etc.).
	Interaction with other road users (safety)	Driving tasks that require anticipation, prediction, and decision-making in relation to the behavior of other road users, particularly in safety-critical situations (absence/presence of other road user's, prediction of other road user's behavior).
	In-vehicle control	Driving-related tasks involving interaction with in-vehicle systems or interfaces that may occur concurrently with vehicle operation.

Table 3. *Overview of possible manipulations linked to peripheral vision factor(s).*

Manipulation type (PV Factor)	Category	Manipulation Type	Examples	Combination
Perception	Eccentricity	Varying the angular distance of stimuli (e.g., hazard, target or traffic lights)	Stimuli placed at 3, 10, or 20 degrees.	-
Perception	Field of view limitation	Manipulating visible field width	Simulator displays restricted the full-width of the screens	-
Perception	Peripheral information degradation	Reducing quality or removing cues from road edges or periphery	Degraded road edges or background masking	-
Perception	Moving window or gaze-contingent restriction	Dynamically controlling which regions are visible based on gaze	Gaze-contingent mask during simulated driving	-
Cognition	Cognitive load (without visual or auditory load)	Working memory, auditory n-back, (imagery) verbal task	Counting backwards, cell-phone conversation	-
Cognition	Visual load	Secondary visual task in-vehicle or display task	Map reading or in-car visual search	-
Cognition	Auditory load (conversation)	Passenger conversation, listening to music or cognitive conversation task	Dialogue with a passenger or answering trivia questions	-
Cognition	Predictability (cueing and expectancy)	Exogenous and endogenous cues	Exogenous cueing (e.g., brake light, flashing billboard) and endogenous cueing (a road sign warns you for a hazard later)	-
Cognition	Emotional or arousal state	Induced anger, anxiety or stress	Negative feedback, stress-inducing time pressure or emotionally auditory stimuli	-
Combination	Environmental complexity	Traffic density, presence of pedestrians or cyclists	High vs flow traffic, unexpected hazards	Perception and cognition
Combination	Motor dual-task	Secondary motor task	Manual button pressing task or touchscreen (navigation) task	Cognition and motor
Combination	Speed or time pressure	Faster driving or temporal constraints	Increased driving speed, reduced time-to-collision, shortened response windows	Perception and cognition
Other	Expertise	Skill level as a group	Learner driver’s vs licensed drivers	-
Other	Experience (familiarity)	Driving experience	Comparison of novice vs experienced drivers	-

Dynamic Advertising on Vehicles: Effects on Driver Glance Behavior

Tibor Petzoldt [1*], Lisa Zwicker [1,2]

[1] TU Dresden, Chair of Traffic and Transportation Psychology, Hettnerstr. 1-3, 01069 Dresden, Germany
[2] TÜV | DEKRA arge tp 21 GmbH, Wintergartenst. 4, 01307 Dresden, Germany
[*] tibor.petzoldt@tu-dresden.de

Abstract: Vehicle-mounted advertising is becoming increasingly feasible due to advances in digitalization and illumination technologies. Unlike roadside advertisements, mobile ads can remain in a driver's field of view for extended periods, potentially increasing attentional demands and distraction. Despite growing industry interest, little is known about how temporal characteristics and content relevance of vehicle-based advertising affect drivers' glance behavior. We conducted a driving simulator study with 42 participants in highway and urban scenarios. A preceding truck displayed rear-mounted advertisements that varied in temporal variability (static, alternating, dynamic) and relevance. Driver gaze was recorded using a head-mounted eye-tracking system, and analyses focused on fixations directed at the truck over 6- and 15-second windows. Results showed that relevant content increased fixation frequency and duration, and alternating or dynamic advertisements further amplified attention toward the truck. Accumulated fixation durations often exceeded critical thresholds, particularly for variable content. Temporal variability strongly influenced gaze behavior, with relevance further modulating attentional focus. These findings demonstrate that vehicle-based advertising can increase driver distraction. They provide a foundation for further research and for considering technical or regulatory measures to mitigate attentional risks associated with this emerging advertising medium.

1. Introduction

Stationary advertising media such as backlit city-light boards, screens, and posters are well established, whereas mobile advertising is largely limited to vinyl-wrapped sections of public and, less commonly, private vehicles. Growing demand for innovation in the advertising industry is expected to drive the exploration of previously unused or underutilized spaces (Siegert & Brecheis, 2017), increasing the presence of advertising on vehicles. At the same time, advances in digitalization and lighting technology may enable more feasible and cost-effective illuminated and video-based vehicle advertising (Alt, Evers & Schmidt, 2009; Siegert & Brecheis, 2017). Consequently, vehicle-based advertising is expected to become more variable, ranging from static images to alternating messages and fully dynamic displays.

A key difference between roadside and mobile advertisements is exposure duration. Roadside ads are visible only briefly as vehicles pass, whereas mobile billboards can remain in a driver's field of view for much longer. While this creates opportunities for dynamic content from an advertising perspective, prolonged exposure also increases distraction risk. It may also alter the mechanisms of attentional capture (see the SEEV model; Wickens & Horrey, 2008). For roadside advertisements, bottom-up factors are likely to dominate. Their *salience* and spatial placement relative to drivers' typical gaze direction, that is, the *effort* required to attend to them, largely determine the attention they attract (e.g., Wilson & Casper, 2016). With prolonged exposure, particularly when the advertisement remains near the driver's central field of view, top-down influences become more relevant. Factors such as the frequency of message changes, shaping the *expectation* that relevant information may appear, and the informational *value* or relevance of the content may therefore play a larger role in attention allocation and, consequently, in driver distraction and performance.

These considerations highlight the need to understand how mobile advertisements influence drivers' gaze behavior, particularly when exposure is prolonged, for example when following an advertising-bearing vehicle. The present study experimentally examined how two characteristics of vehicle-mounted advertising, (a) the temporal variability of the content (static, alternating, dynamic) and (b) the relevance of the displayed information, affect drivers' glance behaviour. (More detailed information on the present study, as well as additional results from the project, can be found in the project report; see Zwicker, Schulze, & Petzoldt, 2025.)

2. Method

2.1 Driving Simulation

A driving simulator study was conducted to address the research question. The experiment was implemented using the SILAB driving simulation software (Würzburger Institut für Verkehrswissenschaften GmbH, WIVW), which provided a 180° horizontal field of view. Driver gaze behavior was recorded using a head-mounted eye-tracking system (iView X, Version 2.8; SensoMotoric Instruments GmbH, SMI).

The study included both highway and urban driving scenarios. In total, 26 scenarios were implemented, of which 15 are relevant for the analyses reported here. In the highway scenarios (see Fig. 1), participants drove on a multi-lane highway with a speed limit of 100 km/h and encountered a preceding vehicle displaying rear-mounted advertising. Overtaking opportunities were manipulated by varying traffic density on the left lane, which was either congested and prevented overtaking or clear and allowed participants to pass the advertising vehicle. Only the non-overtaking scenarios are considered in the present analysis. In addition,



Fig. 1. *Example screenshot out of the driving simulation; here: highway scenario*

participants encountered situations in which two trucks travelled side-by-side during an overtaking maneuver, with the advertisement displayed on the left vehicle.

In the urban scenarios, participants drove in a city environment with the standard German inner-city speed limit of 50 km/h. They encountered the advertising vehicle while waiting at a traffic light. Overtaking was not possible in these scenarios due to the street layout and traffic density.

2.2 Experimental Manipulation

Advertisements were presented on the back of a truck. To avoid emotional or opinion-based bias, no real products or brands were used. Stimuli consisted of images of animals, specifically monkeys or big cats (see Fig. 1 for an example). Relevance was manipulated by informing participants that they would be questioned about monkeys in the post-drive questionnaire, making monkeys relevant and big cats non-relevant.

Advertisements were either static, alternating, or dynamic. Static ads displayed a single unchanging image throughout the scenario. Alternating ads rotated between two images every 5 seconds. Dynamic ads consisted of 5-second looping video sequences.

2.3 Participants

The final sample comprised 42 participants (40% female, 60% male). Twenty-one participants were aged 30 years or younger, 13 were aged 31-40 years, 5 were aged 41-50 years, and 3 were aged 51-60 years. Eligibility criteria required participants to be no older than 60 years and to have held a valid driver's license for at least three years. Participants received 20 euros for their participation.

Due to data quality issues, primarily related to unsuitable glasses that interfered with eye-tracking measurements, usable glance data were available for 36 participants. In addition, glance data were not available for all experimental conditions for all participants because of technical recording issues or deviations from the driving instructions (e.g., overtaking the lead vehicle).

2.4 Data Analysis

Analyses focused on fixations directed at the preceding truck, characterized by their number, average duration, and accumulated duration over 6- and 15-second observation windows, following the thresholds suggested by Klauer et al. (2010), where a glance away from the road is considered critical if it exceeds 2 seconds within 6 seconds or 3 seconds within 15 seconds. It should be noted, however, that in this study fixations on the truck cannot be straightforwardly classified as traffic-relevant or irrelevant. Participants may have been attending to the vehicle as a traffic object, its speed, or its position, as well as to the advertising displayed. Nevertheless, comparing conditions is informative because it indicates whether participants fixated more frequently or for longer periods under specific conditions.

3. Results

The eye-tracking data generally aligned with expectations. Significant effects were most evident in the urban scenarios, where relevant advertisements led to a comparatively "calmer" gaze pattern: participants made fewer fixations but each fixation lasted longer.

In urban scenarios, trucks displaying relevant content (monkeys) were fixated more frequently and for longer durations than those showing non-relevant content (big cats). Accumulated fixation durations in the 6-second window exceeded the 2-second threshold for alternating and dynamic stimuli (approximately 2.34 s and 2.44 s, respectively), while static images remained below it (1.74 s). In the 15-second window, all conditions exceeded the 3-second threshold, with fixations lasting approximately 5.32 s for static, 6.32 s for alternating, and 6.19 s for dynamic stimuli.

Temporal variability of the advertisement also influenced gaze behavior. In urban scenarios, alternating stimuli led to more fixations (mean 6.6) than static stimuli (mean 5.4) over the 6-second window, and both alternating and dynamic content were associated with longer accumulated fixation durations than static images.

Similar patterns were observed on the highway during the prolonged truck overtaking maneuver. Dynamically changing advertisements resulted in longer fixation durations than static images (mean and accumulated fixations over 15 seconds). All conditions exceeded the 3-second threshold in the 15-second window, but the accumulated fixation duration varied markedly: 3.03 s for static, 5.28 s for alternating, and 5.54 s for dynamic content.

4. Discussion & Conclusions

The present study provides evidence that (digital) advertising on vehicles carries potential safety risks for road users. Eye-tracking results showed that both relevance and temporal variability of rear-mounted advertisements significantly influenced drivers' gaze behavior. Relevant content increased fixation frequency and duration, while alternating or dynamic content further amplified attention allocation compared with static images. These findings align with participants' subjective assessments in post-drive questionnaires (not reported here), which highlighted concerns about distraction and safety when following advertising-bearing vehicles.

Both content relevance and variability are thus key factors in the attentional impact of vehicle-based advertising. While the relevance of displayed information cannot easily be regulated, the temporal variability could be addressed technically. Measures could include prohibiting continuously

dynamic, video-like content or enforcing minimum display durations for static content. However, the study also indicates possible additive effects: multiple advertising surfaces with changing content in the driver's field of view can increase the frequency of attention-demanding changes. Unlike fixed roadside advertisements, where the number and placement of displays can be controlled, it is technically challenging to regulate the simultaneous presence and interactions of multiple advertising vehicles on the road.

Although compensatory behaviors by drivers may prevent severe incidents in most real-world encounters, the results do not imply that vehicle-based advertising is harmless. Under certain conditions, such as complex traffic layouts, multiple advertising vehicles, or adverse weather, cumulative attentional demands could create a worst-case scenario. Subjective workload ratings further suggest that prolonged exposure may contribute to driver fatigue, which in turn increases accident risk.

Taken together, these findings underscore the need for a societal discussion regarding vehicle-mounted advertising. The potential commercial benefits of this new advertising medium must be weighed against documented attentional impacts and likely safety consequences. Technical and regulatory measures targeting temporal variability may mitigate risk, but inherent limitations in controlling exposure highlight the importance of careful evaluation before widespread implementation.

5. Acknowledgments

This report is based on parts of the research project carried out at the request of the German Federal Ministry for Transport, represented by the German Federal Highway Research Institute, under research project No. 82.0742/2019. The authors are solely responsible for the content.

During the preparation of this work, the authors used ChatGPT in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Alt, F., Evers, C., & Schmidt, A. (2009, May). Users' view on context-sensitive car advertisements. In *International Conference on Pervasive Computing* (pp. 9-16). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-01516-8_2

Klauer, S. G., Guo, F., Sudweeks, J. & Dingus, T. A. (2010). *An Analysis of Driver Inattention Using a Case-Crossover Approach on 100-Car Data: Final Report.* Blacksburg: Virginia Tech Transportation Institute.

Siegert, G. & Brecheis, D. (2017). Werbung in der Medien- und Informationsgesellschaft. Wies-baden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15885-9

Wickens, C. D., & Horrey, W. J. (2008). Models of attention, distraction, and highway hazard avoidance. *Driver distraction. Theory, effects and mitigation*, 249-279.

Wilson, R. T., & Casper, J. (2016). The role of location and visual saliency in capturing attention to outdoor advertising: How location attributes increase the likelihood for a driver to notice a billboard ad. *Journal of Advertising Research, 56*(3), 259-273. https://doi.org/10.2501/JAR-2016-020

Zwicker, L., Schulze, C., & Petzoldt, T. (2025). *Werbung an Fahrzeugen* (Berichte der Bundesanstalt für Straßen- und Verkehrswesen, Reihe F: Fahrzeugtechnik, Heft 169). Fachverlag NW in der Carl Ed. Schünemann KG. https://doi.org/10.60850/bericht-f169

Reducing inattentional blindness to warning signals: the influence of meaning, emotional value, and task context

Katarina Pavic[1,2], Barbara Vieira Ribeiro[1], Laetitia Silvert[1]

[1] Université Clermont Auvergne, CNRS, Laboratoire de Psychologie Sociale et Cognitive, LAPSCO, F-63000 Clermont-Ferrand, France
[2]Nantes Université, Univ Angers, Laboratoire de psychologie des Pays de la Loire, LPPL, UR 4638, F-44000 Nantes, France

Abstract: Camera Monitoring Systems can improve truck drivers' visibility but also increase the amount of visual information they must process, potentially making them more vulnerable to inattentional blindness (IB), i.e., the failure to notice an unexpected object in plain sight. Optimizing warning signals is therefore critical to ensure that they are effectively perceived and to mitigate the potentially severe consequences of IB in driving contexts. Across two experiments involving more than 400 participants, we examined whether stimuli with semantic or emotional value are less susceptible to IB than visually matched neutral, meaningless stimuli, and whether pre-activating them by manipulating the task context could further enhance awareness. In Study 1, an unexpected stimulus appeared while participants performed a line-length judgment task. Neither the semantic nor emotional value of the stimulus reduced IB in that case. In Study 2, the same task was embedded in a driving scenario expected to pre-activate meaningful and emotional stimuli. In this context, the results revealed increased localization rates for the emotional stimulus only. Overall, these findings suggest that semantic or emotional value alone does not ensure detection but may facilitate localization when aligned with the task context.

1. Introduction

Camera monitoring systems (CMS) in trucks use external cameras and in-cab displays to replace or complement traditional mirrors, giving the driver improved visibility around the vehicle. In order to help drivers detect hazards faster, the video feed can be digitally enhanced with visual overlay to indicate obstacles or vulnerable road users through object-detection technologies. While CMS undoubtedly enhance visibility, they also add to the visual information the driver must process, which could lead them to overlook some alerts. Indeed, excessive increases in perceptual load, by reducing the resources available for processing unexpected events, may render drivers particularly vulnerable to Inattentional Blindness (IB; Matias et al., 2021).

IB refers to the failure to consciously perceive a visible but unexpected stimulus when attention is engaged in another task (Mack & Rock, 1998). Although common in everyday life, IB can lead to dangerous and even tragic consequences in driving contexts (Hills, 1980; Pammer et al., 2015, 2017). Identifying which stimulus characteristics influence IB is therefore crucial for designing warning signals that are less likely to be missed.

Pioneering studies demonstrated that IB rates decrease when unexpected stimuli (US) carry particular significance for individuals (the *meaningfulness−signal hypothesis*; Mack & Rock, 1998). From there, the semantic/emotional value of the unexpected stimulus has been considered as one of the main determinants of attention capture. Surprisingly, however, this possibility has been relatively underexplored, and the available evidence remains inconsistent. Several factors may account for these discrepancies. First, experimental protocols vary widely across studies, making comparisons difficult (Redlich et al., 2020). Second, the operationalization of "meaningfulness" differs substantially across investigations. For instance, some studies have used participants' own names (Mack & Rock, 1998), whereas others have relied on neutral or expressive face pictograms (e.g., Redlich et al., 2021) or threatening animals (spiders; New & German, 2015). Third, differences in IB rates for different US may sometimes arise from variations in low-level visual characteristics between stimuli, rather than from their semantic or emotional relevance (Redlich et al., 2021). As a result, it is difficult to disentangle the respective contributions of semantic meaningfulness and emotional salience.

Another factor that can affect the detectability of the US is the observer's attentional set regarding the primary task: US are more likely to be perceived when they are similar to the objects that are being attended as part of the primary task. For instance, participants tracking black items detect unexpected black objects more often than white ones (Most et al., 2001). Here, US may reach conscious perception through top-down additional activation resulting directly from their task-relevance (Dehaene et al., 2006). If so, an US should also be detected more frequently when one of its features is activated prior to the primary task. Previous research suggests that pre-activating a physical feature of an unexpected stimulus (e.g., its colour) does not influence inattentional blindness rates (Kreitz et al., 2015). However, to our knowledge, the effect of pre-activating the whole stimulus' concept has not yet been investigated.

Thus, the aim of the present studies was to investigate whether and when unexpected visual stimuli with high semantic value and emotional value may capture attention more effectively than a visually matched neutral, meaningless stimulus. The aim of Study 1 was to examine whether semantic and emotional properties of the US alone reduce IB. In Study 2, we manipulated the participants' attentional set by embedding the task into a driving scenario. We expected

this manipulation to boost the perception of the meaningful US.

2. Study 1

2.1 Method

2.1.1 Participants

The sample consisted of 246 French-speaking participants (M = 33.3, SD = 10.0 years old) recruited via the Prolific platform.

2.1.2. Stimuli

We designed three visually matched stimuli: a scrambled shape, a stick figure, and a danger icon. Their semantic and emotional values are presented in **Figure 1**.

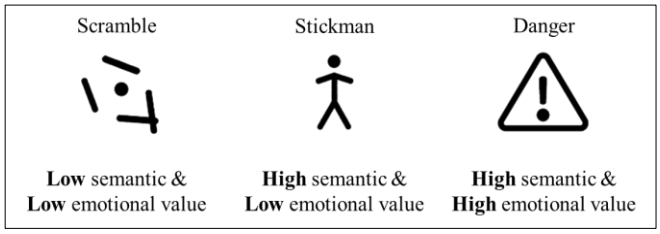


Figure 1*. Illustration of the stimuli and their semantic and emotional values. These values were validated in a pretest conducted on an independent sample of 60 participants.*

2.1.2 Procedure

Participants completed a static IB task (see **Figure 2**) adapted from Newby & Rock (1998). On each trial, they judged which arm of a briefly presented central cross was longer. After two practice trials and four test trials, an unexpected stimulus (US) appeared in the fifth critical trial. The US consisted of one of the three stimuli designed for this study and appeared without warning in a quadrant of the screen. Participants reported whether they *detected* the US, indicated its *location*, and *identified* it using a forced-choice response. They also rated their confidence on a 4-point Likert scale (0 = no confidence, 3 = high confidence).

The final three trials were control trials. In the biased-attention trial, no US appeared, whereas the US was shown in the divided- and full-attention trials, with the latter instructing participants to focus solely on the stimulus. Detection, localisation, and identification responses were required for all three trials.

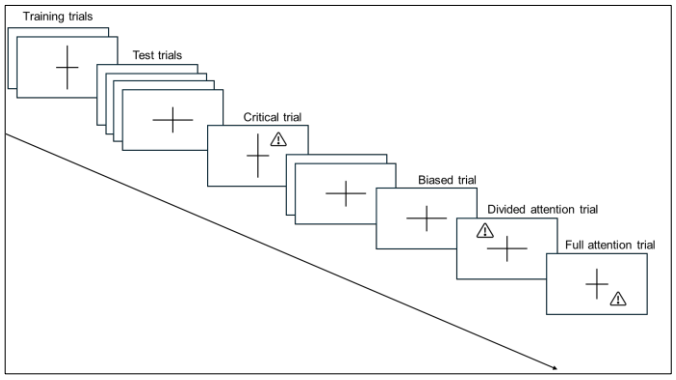


Figure 2. *Schematic illustration of the static IB task. Note: The unexpected stimulus is not shown to scale.*

2.2 Results

Binary logistic regressions were conducted on detection, localisation, and identification rates of the US during the critical trial (see **Figure 3**)**.** A significant effect of the US was found on identification rates (p =.03). However, contrary to our expectations, identification was lowest for the stickman US (OR = 0.38, 95% CI [0.18, 0.77], p =.009). No effect of the US was found on detection (p =.21) or localisation (p =.22).

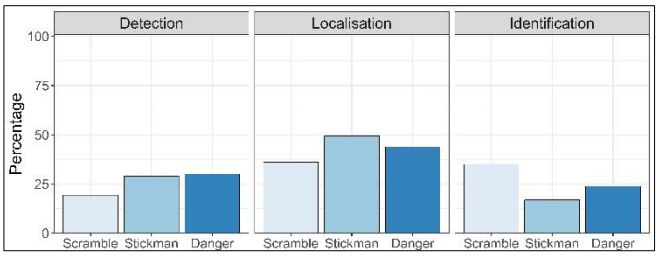


Figure 3. *Percentage of participants in Study 1 who correctly detected, localised, and identified the unexpected stimulus presented during the critical trial.*

3. Study 2

3.1 Method

3.1.1 Participants

The sample consisted of 243 French-speaking participants (M = 32.9, SD = 9.5 years old) recruited via the Prolific platform.

3.1.2 Procedure

The same stimuli and procedure as in Study 1 were used. However, the task instructions were modified to embed the IB task in a driving scenario. Participants were first presented with a brief introductory text that framed the experiment in a road-traffic context. Specifically, the two lines used in the length-judgment task were described as representing a road intersection.

3.2 Results

Binary logistic regressions were again conducted on detection, localisation, and identification rates of the US during the critical trial (see **Figure 4**). A significant effect of the US was found on localisation (p =.01) and identification rates (p =.02). As expected, localisation performance was highest for the danger icon (OR = 2.82, 95% CI [1.50, 5.41], p = .001). However, contrary to our expectations, identification performance was lowest for the stickman (OR = 0.438, 95% CI [0.201, 0.919], p = .03). No effect of the US was found on detection (p = .13).

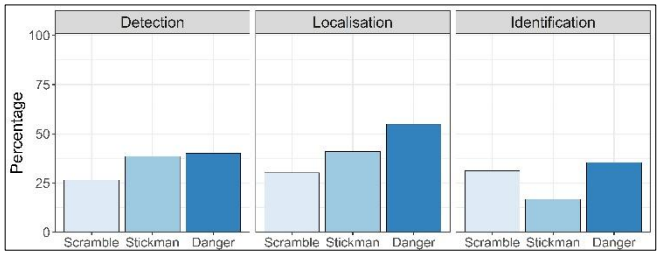


Figure 4. *Percentage of participants in Study 2 who correctly detected, localised, and identified the unexpected stimulus presented during the critical trial.*

4. Additional methodological checks

In both studies, we observed high detection, localization, and identification rates in divided- and full-attention trials, indicating that the results cannot be attributed to perceptual limitations of the US.

Moreover, participants' ratings of the arousal, valence, and meaning of the three stimuli replicated the pre-test results, confirming that the stimuli's semantic and emotional values were perceived as intended.

5. Discussion

Taken together, the findings of Study 1 indicate that neither the semantic nor emotional value of US influenced IB in a decontextualized task. Although unexpected, this result is consistent with the contingent capture hypothesis (Folk & Remington, 1992), suggesting that the US failed to capture attention because they did not match participants' attentional set (i.e., judging line lengths).

Study 2 provides a more nuanced perspective: when the IB task was embedded in a driving scenario designed to pre-activate meaningful stimuli, their detection and identification did not improve. However, localisation rates increased for the emotional stimulus alone (the danger icon resembling a road sign).

Taken together, these results offer novel theoretical and practical contributions that will be discussed more thoroughly during the presentation. From a theoretical perspective, they challenge the role of semantic and emotional salience in attentional capture, which is commonly reported in the literature. From a practical perspective, they suggest that implementing visual warning signals with both semantic and emotional value may enhance their localisation in driving contexts.

6. Acknowledgments

This research is part of the DOLPHIN project (led by Renault Trucks) which has received funding from the French government as part of France 2030 and the European Union – Next Generation EU as part of the France Relance plan.

The authors would like to express their gratitude to Hélène Fantinutti (RT), David Chandon (RT), and Thibaut Jean-Pierre (Gauzy) for their assistance in this project as well as to the participants of these studies.

References

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of experimental psychology. Human perception and performance*, *18*(4), 1030–1044.

Hills, B. L. (1980). Vision, Visibility, and Perception in Driving. *Perception*, *9*(2), 183-216. https://doi.org/10.1068/p090183

Koivisto, M., & Revonsuo, A. (2007). How meaning shapes seeing. Psychological Science, 18(10), 845–849. https://doi.org/10.1111/j.1467-9280.2007.01989.x

Kreitz, C., Schnuerch, R., Furley, P. A., Gibbons, H., & Memmert, D. (2015). Does semantic preactivation reduce inattentional blindness? Attention, Perception & Psychophysics, 77(3), 759–767. https://doi.org/10.3758/s13414-014-0819-8

Mack, A., & Rock, I. (1998). Inattentional Blindness : Perception without Attention. In *Visual Attention*.

Matias, J., Belletier, C., Izaute, M., Lutz, M., & Silvert, L. (2022). The role of perceptual and cognitive load on inattentional blindness: a systematic review and three meta-analyses. Quarterly Journal of Experimental Psychology, 75, 1844-75. https://doi.org/10.1177/17470218211064903

New, J. J., & German, T. C. (2015). Spiders at the cocktail party: An ancestral threat that surmounts inattentional blindness. *Evolution and Human Behavior*, *36*(3), 165-173.

Newby, E. A., & Rock, I. (1998). Inattentional Blindness as a Function of Proximity to the Focus of Attention. *Perception*, *27*(9), 1025-1040. https://doi.org/10.1068/p271025

Pammer, K., Bairnsfather, J., Burns, J., & Hellsing, A. (2015). Not All Hazards are Created Equal : The Significance of Hazards in Inattentional Blindness for Static Driving Scenes. *Applied Cognitive Psychology*, *29*(5), 782-788. https://doi.org/10.1002/acp.3153

Pammer, K., Sabadas, S., & Lentern, S. (2017). Allocating Attention to Detect Motorcycles : The Role of Inattentional Blindness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60*(1), 5-19. https://doi.org/10.1177/0018720817733901

Redlich, D., Memmert, D., & Kreitz, C. (2021). A systematic overview of methods, their limitations, and their opportunities to investigate inattentional blindness. *Applied cognitive psychology*, *35*(1), 136-147.

Redlich, D., Memmert, D., & Kreitz, C. (2022). Clarifying the effect of facial emotional expression on inattentional blindness. *Consciousness and cognition*, *100*, 103304.

Braking without looking: understanding driver responses to sudden lead-vehicle deceleration using a forced peripheral vision paradigm

Malin Svärd[1,*]

[1] Volvo Cars Safety Centre, Volvo Cars, 418 78 Göteborg, Sweden
[*] malin.svard@volvocars.com

Abstract: Understanding how drivers use peripheral vision information to guide responses in safety-critical situations is essential to develop realistic driver behavior models and improve the accuracy of virtual safety evaluations. This study investigates how drivers' brake response times (BRTs) are influenced by gaze eccentricity angles, gaze-manipulation method, and expectancy. In a driving simulator experiment, participants were exposed to sudden lead-vehicle decelerations while their gaze direction was manipulated using a secondary-task-based method or forced peripheral vision. The BRTs increased with eccentricity even when foveal cues were severely restricted, highlighting a degradation in peripheral-vision processing. Both gaze-manipulation methods resulted in similar BRTs but differed in compliance and likelihood of braking. The large BRT difference between expected and unexpected braking cautions against using expected-condition data for assessing safety-critical driving behavior and for ADAS design.

1. Introduction

Visual attention to the road ahead is essential for safe driving. Yet in-vehicle displays and nomadic devices frequently draw the driver's gaze away from the roadway, raising concerns about how off-road glances impact driver reactions safety-critical events. Prior research has primarily treated attention as an all-or-nothing phenomenon where the driver is either fully attentive or completely inattentive. This perspective has shaped regulations aimed at limiting off-road glance frequency and duration to improve road safety.

Drivers can often rely on peripheral vision to control the vehicle, but the degree to which peripheral vision guides timely reactions to critical situations remains unclear. A notable exception is the study by Svärd et al. (2021), where glance and brake reaction times to a sudden lead-vehicle deceleration was examined as function of gaze eccentricity. They found that glance response times were unaffected by eccentricity, whereas the time interval between gaze redirection and brake initiation increased with larger gaze angles. The study faced challenges in motivating sufficiently long off-road glances via a secondary task, resulting in many participants being excluded because they looked back for reasons unrelated to the stimulus from the lead vehicle.

Secondary tasks are commonly used to manipulate gaze behavior, but alternative methods exist. The "forced peripheral driving paradigm", introduced by Summala et al. (1996), was conceptually inspired by early work showing that drivers can use information from both foveal and extra-foveal vision (Bhise & Rockwell, 1971). Here, we refer to it as the *forced peripheral vision paradigm* to emphasize its visual manipulation aspect. In this paradigm, drivers are instructed to maintain gaze at controlled eccentric positions, often corresponding to in-vehicle display locations. Summala and colleagues applied the method to study lane keeping (Summala et al., 1996) and detection of lead-vehicle longitudinal deceleration (Lamble et al., 1999; Summala et al., 1998), reporting degraded control performance with increasing eccentricity. However, the scenarios were low-speed and non-critical, and compliance with gaze instructions was not fully disclosed.

In the present study, we apply the forced peripheral paradigm to a critical longitudinal situation. We investigate whether drivers can initiate timely braking using peripheral vision alone and how performance varies with eccentricity. By comparing the results to prior findings from a secondary-task-based study, we assess the suitability of the forced peripheral paradigm in driver behavior research.

2. Method

The data analyzed here originate from the same driving-simulator experiment as described in Svärd et al. (2021). In the study, participants followed a lead-vehicle on a divided highway and were exposed to sudden decelerations requiring rapid driver responses. Here, we analyze a previously unpublished part of the experiment, where the forced peripheral vision paradigm was applied. Here, we explicitly instructed the participants that the lead vehicle would brake while they were looking away. These trials occurred immediately after the secondary-task-based segments of the experiment.

Participants were assigned into three groups, each exposed to three lead-vehicle braking trials while fixating their gaze on secondary tasks presented at predefined eccentricities (12°, 40°, and 60°). Only the order of eccentricities differed between groups. All participants were instructed to maintain gaze fixation on the secondary-task display and to brake when necessary, without first redirecting gaze toward the road ahead. Brake response time (BRT), gaze behavior and vehicle kinematics were continuously recorded.

Brake response time was defined as the interval between lead-vehicle deceleration onset and the first increase in brake-pedal pressure. Gaze direction was tracked to confirm compliance with instructions. Trials in which the participants redirected their gaze toward the road before braking were excluded from the main analysis but examined separately. We also performed a separate analysis on the number of trials in which participants failed to react before

collision. The results were compared the secondary-task-based gaze manipulation method used in the studies by Svärd et al., (2021, 2022), including both an unexpected and an expected lead-vehicle brake event. This comparison enables us to assess whether the forced peripheral vision paradigm offers stronger control of gaze behavior while still yielding reliable estimates of peripheral-vision-based reaction times.

3. Results

Driver responses were analyzed by comparing BRTs across eccentricities and glance-manipulation methods, investigating the effects of gaze manipulation and driver expectancy, and analyzing response types.

3.1 BRTs with forced peripheral vision

The BRTs were aggregated per eccentricity angle for all three groups, resulting in 106 data points (45 at 12°, 42 at 40°, and 19 at 60°). We found statistically significant group differences (ANOVA: $F=11.23$, $p<0.001$; Kruskal–Wallis: $H=19.52$, $p<0.001$), with longer BRTs at higher eccentricity angles (Fig. 1). Pairwise Mann Whitney comparisons with Holm correction for multiple testing confirmed that all groups differed from each other with moderate to large effects ($p<0.010$, $r=0.32$-0.62).

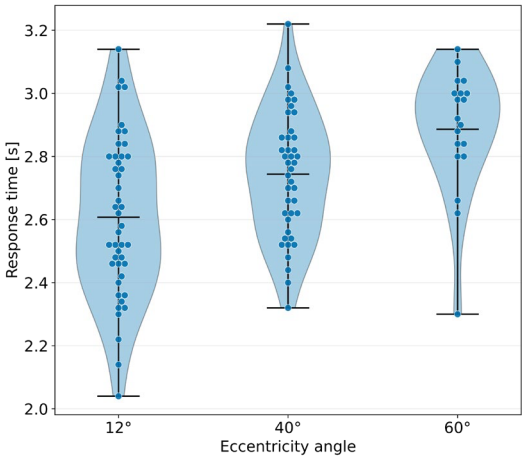


Fig. 1 *Violin plots of BRTs for each eccentricity. The markers represent individual driver responses.*

3.2 Impacts from glance manipulation method choice and expectancy

We compared the BRTs from the second forced peripheral vision trial to the BRTs from the last, expected, brake event using secondary-task gaze manipulation (Fig. 2). The eccentricity angles were matched between the groups, ensuring that each group was exposed to the same eccentricity in both conditions. No significant BRT differences were found between the conditions ($p\geq0.45$; $r=0.18$–0.50). Further, the forced peripheral vision condition was compared to the first, unexpected, brake scenario. All groups showed substantially longer BRTs in under the latter condition, but statistical significance ($p<0.001$; $r=0.71$–0.76) was only reached for the lower eccentricity angles (<60°).

3.3 Missed braking and non-compliance rates

Differences in non-compliance and missed-braking rates were compared between forced peripheral vision and secondary-task gaze manipulation (last, expected brake event) conditions (Fig. 3). We observed a significant reduction in non-compliance rates across all groups for the forced peripheral vision condition (Cochran-Mantal-Haenszel odds ratio=0.29; $p<0.001$), at the cost of more frequent missed braking cases for high eccentricity angles (non-significant across groups).

4. Discussion

Prior research suggests that different mechanisms drive gaze redirection and brake response times, with only the latter being dependent on eccentricity angle (Svärd et al., 2021). Consistently, we here show that drivers BRTs increase with eccentricity also in absence of gaze redirection. Drivers can collect evidence for braking through the peripheral vision system, though the evidence degrades with increased gaze eccentricity. These insights can guide the development of driver-monitoring-system (DMS) based advanced driver assistance systems (ADAS) and contribute to realistic computational driver behavior models for virtual safety assessments.

BRT distributions were comparable across gaze manipulation conditions for all angles when a lead-vehicle brake was expected. The main difference between

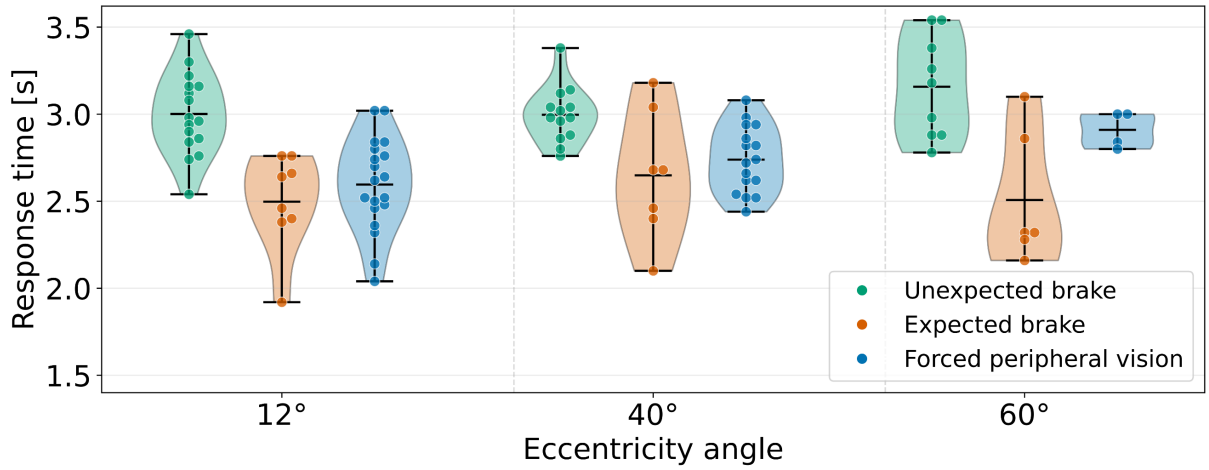


Fig. 2 *Violin plots of BRTs for each eccentricity and experimental condition. The markers represent individual driver responses, while the colors correspond to experimental conditions.*

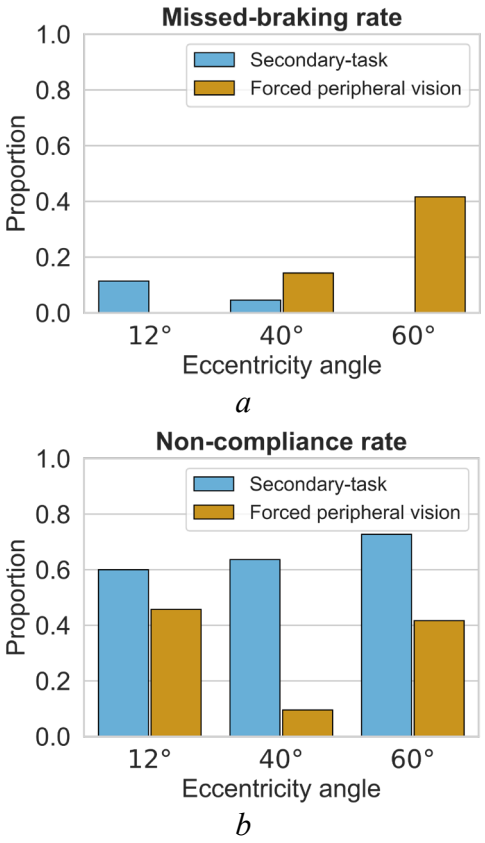


Fig. 3 *Missed-braking and non-compliance rates for the examined gaze-manipulation methods.*
(a) *Missed-braking rates,* ***(b)*** *Non-compliance rates.*

conditions was instead found in participant compliance rates and whether braking was initiated at all. While drivers were more compliant with the forced peripheral vision paradigm, some individuals seemed more hesitant to brake.

The study confirms previous research on repeated brake exposures with a substantial BRT difference between the first (unexpected) and subsequent (expected) brake events (Benderius et al., 2014). This behavioral discrepancy raises questions about the ecological validity of using response times acquired in expected braking conditions to develop critical-event response models or inform ADAS design.

Note that care must be taken when interpreting the absolute BRTs from this study, since all driving was performed in a controlled driving simulator environment which may not be directly comparable with real-world conditions.

5. Conclusions

This study shows that BRTs increase with increasing gaze eccentricity angle, even when foveal cures are severely restricted. Confirming a degradation in driver's ability to access and process visual information at higher eccentricities, the findings emphasize the importance of peripheral vision in driving.

In expected-brake scenarios, forced peripheral vision resulted in BRTs comparable to those obtained with a traditional gaze-manipulation method. However, methodological choice substantially influenced compliance and likelihood of braking. The pronounced BRT gap between expected and unexpected braking events further highlights the risk of using expected-condition data for modelling driver responses or informing ADAS design. Future work should further examine how methodological choices impact the ecological validity of driver-behavior studies.

6. Acknowledgments

The driving simulator experiment was founded by Vinnova, the Swedish governmental agency for innovation, through the strategic vehicle research and innovation program FFI. We want to acknowledge Trent Victor and Jonas Bärgman for conceptualizing this study, and thank Fredrik Granum and Pär Kåbäck for assisting in the data collection process.

References

Benderius, O., Markkula, G., Wolff, K., & Wahde, M. (2014). Driver behaviour in unexpected critical events and in repeated exposures – A comparison. *European Transport Research Review*, *6*(1), 51–60. https://doi.org/10.1007/s12544-013-0108-y

Bhise, V. D., & Rockwell, T. H. (1971). Role of peripheral vision and time-sharing in driving. *Proceedings: American Association for Automotive Medicine Annual Conference*, *15*, 320–341.

Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: implications for positioning of visually demanding in-car displays. *Ergonomics*, *42*(6), 807–815. https://doi.org/10.1080/001401399185306

Summala, H., Lamble, D., & Laakso, M. (1998). Driving experience and perception of the lead car's braking when looking at in-car targets. *Accident Analysis and Prevention*, *30*, 401–407. https://doi.org/10.1016/S0001-4575(98)00005-0

Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *38*(3), 442–451. https://doi.org/10.1518/001872096778701944

Svärd, M., Bärgman, J., Markkula, G., & Ljung Aust, M. (2022). Do Car Drivers Respond Earlier to Close Lateral Motion Than to Looming ? The Importance of Data Selection. *Measuring Behaviour*, 183–191.

Svärd, M., Bärgman, J., & Victor, T. (2021). Detection and response to critical lead vehicle deceleration events with peripheral vision: Glance response times are independent of visual eccentricity. *Accident Analysis and Prevention*, *150*, Article 105853. https://doi.org/10.1016/j.aap.2020.105853

"What Does the Car Know?" Mapping Mental Models of Driver Monitoring Systems

Iris Jestin [1,2*], Catherine Harvey [1], Michael Craven [3], Neil Chadborn[4], Sarah Sharples[5]

[1] Human Factors Research Group, School of Engineering, University of Nottingham, UK
[2] Horizon Centre for Doctoral Training, University of Nottingham, UK
[3] NIHR MindTech Research Centre, Institute of Mental Health, University of Nottingham, UK
[4] NIHR Applied Research Collaboration, University of Nottingham, UK
[5] Faculty of Science and Engineering, University of Manchester, UK
[*] iris.jestin@nottingham.ac.uk

Abstract: With driver monitoring systems (DMS) becoming a required feature in all new vehicles, it becomes important to explore the public's understanding of its working to ensure effective adoption within safe driving systems. This interview study used paper-based visualisation activities alongside a novel methodological approach with AI-generated driving scenario videos as a stimulus to elicit participant mental models of DMS. Analyses identified three archetypes of DMS understanding and four of DMS adoption intention. Findings suggest that a detailed understanding of DMS related technologies does not necessarily translate into a greater intention to adopt them. Instead, acceptability appears to be more strongly shaped by the specific concerns individuals hold about the technology. Future research will investigate how these DMS understanding archetypes potentially influence preferences for different HMI interventions.

1. Introduction

Driver monitoring systems (DMS) use sensors and other means to monitor the driver's state in real-time. This helps to enhance safety by identifying undesirable driver states that may require support, such as distraction, fatigue, and inattention. They are also used to increase user experience by providing measures to enhance driver wellbeing and comfort, such as by changing the lights or music. DMS are soon becoming a required feature in all new vehicles, regardless of their level of vehicle automation (Lenné et al., 2020). As we move towards implementing DMS technologies in vehicles, it becomes crucial to explore how they will be received by wider society (Nees & Liu, 2022). While there exists research on public acceptance, perceptions and attitudes towards DMS (Coyne et al., 2024; Öztürk et al., 2026; Smyth et al., 2021), less is known about the understanding of its operation. The study uses the concept of mental models, as a framework to elicit this understanding of DMS technology.

2. Background

The efficacy of DMS in preventing driver inattention to aid safe interaction with human machine interfaces (HMI) in vehicles is dependent on the accuracy of a driver's mental model of system capabilities and limitations (Merat & Louw, 2020). Mental models are an internal representation of an external reality (Johnson-Laird, 2010). They are simplified descriptions, that humans use in order to understand and predict various socio-technical systems (Andrews et al., 2023). Research that examined a driver's mental model of vehicle automation states that there does not exist a single conceptual model of a system that can be defined as accurate and complete. Hence, the merit of a mental model lies in enabling a user to 'satisfactorily' predict system behaviour (Seppelt, 2020).

Prior research that used an online questionnaire to explore DMS mental models highlighted that drivers were sceptical of DMS having the ability to detect adverse mental states (Nees & Liu, 2022), but believed that most of the DMS monitoring occurred of the driver's interactions with the phone and vehicle console (touch screens, climate controls, music). Acknowledging the lack of publicly available information around DMS functionality, the current study takes a step towards creating clarity by deriving public understanding of this technology. The study, takes inspiration from work that used paper-based visualisation and interview-based mental model building (Kenzie et al., 2024; Revell & Stanton, 2014), to elicit the public's understanding of, and commonly held mental models of DMS operation.

3. Method

3.1 Participants

The 90-minute in-person semi-structured interview study was conducted with 15 participants who held a valid driving license. There was good representation across age groups with the final sample comprising 8 female and 7 male participants, between the ages 22 - 73.

3.2 Study Design and Procedure

Three scenarios were developed using literature, manufacturer information, and relevant regulatory guidance (Euro NCAP, 2024). The first driving scenario represented fatigue detection, chosen because it is an established and familiar DMS capability. The second scenario represented detection of distraction cause by mobile phone use, a recent and less familiar development. The third scenario depicted a futuristic (and consequently unfamiliar) implementation of DMS, in which the vehicle monitors the cognitive and emotional state of an older driver. Presenting scenarios at different levels of technology readiness, and thus familiarity, helped build understanding, as participants started with the most familiar technology and progressed to the least familiar and most futuristic. Scenario videos were created using a novel approach to stimulus generation with Google's AI

video generation tool Veo (see Fig. 1) and Canva for editing. Each video clip ended with a neutral narration prompt such as, "the car intervened to prompt the driver to pay attention to his driving", for scenario 2. To minimise bias, the videos did not show the intervention mode chosen or vehicle automation levels, and the recruitment materials used the term Advanced Vehicle Technologies in place of DMS.

Participants viewed each 50-second video and then answered the same series of interview questions, 'what the car knew about the driver's state', 'how this information was obtained', 'how it informed system responses', and 'what happened to the information after'. Their thoughts were recorded using sticky notes and chart paper. After the three scenarios, participants were guided to create a visual representation of their overall understanding of DMS. The study ended with participants answering a set of DMS acceptability and adoption intention questions.



Fig. 1. *Driving scenario 3 video created on Google Veo AI video generation tool*

4. Analysis and Results

The study generated two types of data for analyses namely; i) participant-created mental model diagrams (see Fig. 2) and their accompanying verbal protocols ii) participants' views on DMS acceptability and intention to adopt it. The data was analysed using a combination of cognitive map analyses (Dixon & Lammi, 2014) and the framework method (Gale et al., 2013). In order to bring a standardisation in the comparison of the participant-created mental model visualisations and better represent the interactions between participant-identified DMS elements, an interpretation of each participant visualisation based on their accompanying verbal protocol was made by the researcher through cognitive maps. This also ensured that a better (aesthetically) represented visualisation would not automatically be seen as a greater understanding of DMS and related technologies.

4.1 DMS Understanding Archetypes

The results generated three archetypes of the participants' detailedness in DMS understanding which are as follows: the Highly Detailed Understanding (HDU), the Moderately Detailed Understanding (MDU) and the Less Detailed Understanding (LDU). The HDU archetype represents users that held a relatively deeper comprehension of how DMS and their related technologies worked, demonstrated by three out of the fifteen participants. Their mental model visualisations were extremely detailed, structured and modular, using relatively accurate terminologies and a greater level of certainty in their explanations. The MDU archetype included four out of the fifteen participants. They demonstrated greater DMS understanding than the LDU archetype but visualised it in lesser detail and a comparatively incomplete and fragmented manner than the HDU participants. Their organisation of DMS knowledge concepts were a combination of linear, circular and modular visualisations, with overlapping traits of both HDU and LDU archetypes. Eight of the fifteen participants were part of the LDU archetype. These participants held inconsistent knowledge concepts of DMS working, as well as a mostly linear, fragmented organisation of these concepts and the interactions between them. They used either incomplete or incorrect terminologies, and their explanations were speculative with a greater degree of uncertainty.

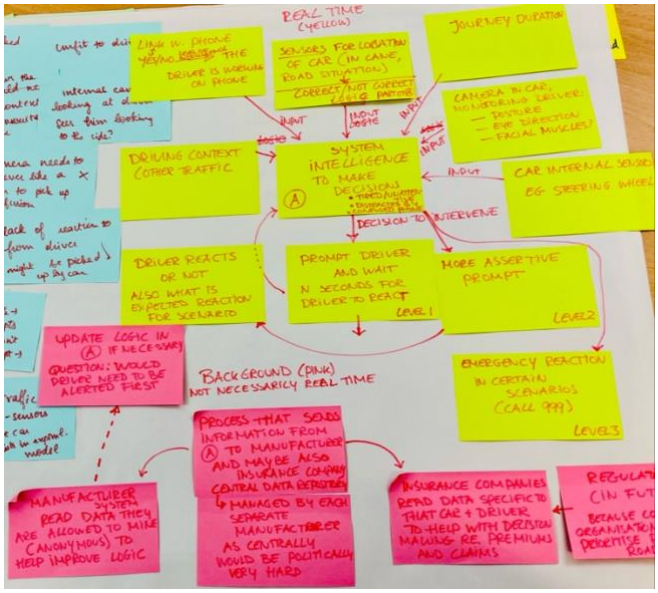


Fig. 2. *Participant 12's mental model visualisation from paper-based activity (HDU archetype)*

The HDU visualisations when compared to the MDU and LDU archetypes, possessed the greatest number of identified DMS elements, the highest number of causal links creating relationships between these elements, and the highest number in the categories/types of causal links between the identified elements. The ability to differentiate between the following three participant-identified DMS elements; 'DMS state detected', 'indicators being monitored to detect said states', as well as 'sensors used to monitor these indicators' were seen to be a distinctive feature to understanding DMS. All participants of the HDU and MDU archetype, when compared to the LDU archetype participants displayed this trait. The LDU archetype participants in comparison, mentioned these three elements either in combination using an umbrella term or used them interchangeably under names such as 'smart technologies data (P5)', 'driver-initiated action (P6)', or simply 'sensors (P7 and P8)', all pointing to some kind of DMS input.

4.2 DMS Acceptability Archetypes

Analysed results produced four archetypes, existing on a spectrum from the least intention to highest intention to



Fig. 3. *Researcher's cognitive map interpretation of visualisation by participant 10 (HDU archetype)*

adopt DMS. The Sceptic archetype (2 of 15 participants) was characterised by a strong disapproval towards the utility or need for DMS. The Reluctant Adopter (3 of 15 participants), possessed an apparent belief in the potential of DMS, but stated various reasons of caution or suspicion which conveyed a lack of confidence. The Pragmatic Realist (5 of 15 participants), while not optimistic about DMS, believed in its need and held a trust in the system to implement it the right way. The Optimistic Enthusiast (5 of 15) displayed qualities of hopeful enthusiasm about the technology's capability, motivated by a strong concern for road safety.

5. Conclusion

Findings revealed that a detailed DMS understanding (HDU archetype) did not necessarily lead to a higher intention to adopt the technology, as none of the HDU participants belonged to the high-adoption Optimistic Enthusiast archetype. Similarly, a less detailed understanding of DMS (LDU archetype) did not correspond with low DMS adoption intention, as none of the participants from this archetype were classified as Sceptics. These results hence suggest that detailed DMS knowledge alone does not drive adoption. Instead, acceptability is potentially more strongly influenced by specific concerns individuals hold about the technology, including 'data sharing beyond the vehicle (P2)', 'big brother monitoring (P3)', 'detection of illegal driver behaviour (P10)', and 'greater perceived vehicle control over the driver (P8)'.

References

Andrews, R. W., Lilly, J. M., Srivastava, D., & Feigh, K. M. (2023). The role of shared mental models in human-AI teams: a theoretical review. *Theoretical Issues in Ergonomics Science*, *24*(2), 129-175. https://doi.org/10.1080/1463922X.2022.2061080

Bobbie Seppelt, T. V. (2020). Driver's Mental Model of Vehicle Automation. In W. J. H. Donald L. Fisher, John D. Lee, Michael A. Regan (Ed.), *Handbook of Human Factors for Automated, Connected, and Intelligent Vehicles* (1st ed., pp. 12). https://doi.org/https://doi.org/10.1201/b21974

Coyne, R., Hanlon, M., Smeaton, A. F., Corcoran, P., & Walsh, J. C. (2024). Understanding drivers' perspectives on the use of driver monitoring systems during automated driving: Findings from a qualitative focus group study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *105*, 321-335. https://doi.org/https://doi.org/10.1016/j.trf.2024.07.015

Dixon, R. A., & Lammi, M. (2014). Cognitive Mapping Techniques: Implications for Research in Engineering and Technology Education. *Journal of Technology Education*. https://doi.org/10.21061/jte.v25i2.a.1

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC medical research methodology*, *13*, 117. https://doi.org/10.1186/1471-2288-13-117

Kenzie, E. S., Wakeland, W., Jetter, A., Lich, K. H., Seater, M., Gunn, R., & Davis, M. M. (2024). Protocol for an interview-based method for mapping mental models using causal-loop diagramming and realist interviewing. *Eval Program Plann*, *103*, 102412. https://doi.org/10.1016/j.evalprogplan.2024.102412

Lenné, M., Roady, T., & Kuo, J. (2020). Driver State Monitoring for Decreased Fitness to Drive. In. https://doi.org/10.1201/b21974

Natasha Merat, Tyron Louw (2020). Allocation of Function to Humans and Automation and the Transfer of Control. In W. J. H. Donald L. Fisher, John D. Lee, Michael A. Regan (Ed.), *Handbook of Human*

Factors for Automated, Connected, and Intelligent Vehicles (1st ed., pp. 19). https://doi.org/10.1201/b21974

Nees, M. A., & Liu, C. (2022). Mental models of driver monitoring systems: Perceptions of monitoring capabilities in an online U.S-based sample. *Transportation Research Part F: Traffic Psychology and Behaviour*, *91*, 484-498. https://doi.org/10.1016/j.trf.2022.10.012

Öztürk, İ., Lehtonen, E., Madigan, R., Lee, Y. M., Aittoniemi, E., & Merat, N. (2026). Decoding acceptance of driver monitoring systems: Evaluating alternative measurement models, cross-country variations, and automated vehicles: Applying the UTAUT framework. *Transportation Research Part F: Traffic Psychology and Behaviour*, *83*, 179-191. https://doi.org/https://doi.org/10.1016/j.trf.2021.10.003

behavioural intention. *Transportation Research Part F: Traffic Psychology and Behaviour*, *116*, 103384. https://doi.org/https://doi.org/10.1016/j.trf.2025.103384

Revell, K. M. A., & Stanton, N. A. (2014). Case studies of mental models in home heat control: Searching for feedback, valve, timer and switch theories. *Applied Ergonomics*, *45*(3), 363-378. https://doi.org/https://doi.org/10.1016/j.apergo.2013.05.001

Smyth, J., Chen, H., Donzella, V., & Woodman, R. (2021). Public acceptance of driver state monitoring for

The European New Car Assessment Programme (2024). Safe driving driver engagement. https://www.euroncap.com/media/83736/euro-ncap-protocol-safe-driving- driver-engagement-v09.pdf.

I2Connect: Intelligent, Interactive, and Connected Next Generation of Real-time Driver Safety Support Systems

Paul Hemeren [1*], Andreas Absér [2], H. Joe Steinhauer[1], Yacine Atif[1], Mikael Lebram[1], Alicia Bernsland[2], Azra Habibovic[2], Henrik Lind [3], Lorenzo Mirante[3], Yuri Tarakanov [4]

[1] University of Skövde, Department of Informatics, Box 408, Skövde, Sweden
[2] Traton Group, Senior Human Factors Specialist, Box 900, Stockholm, Sweden
[3] Smart Eye, Chief Research Officer, Masthamnsgatan 3, Gothenburg, Sweden
[4] Viscando, Business Development, Anders Carlsson gata 14, Gothenburg, Sweden
[*]paul.hemeren@his.se

Abstract:
Navigating intersections is challenging for heavy vehicle drivers due to large dimensions, blind spots, and long stopping distances of their vehicles. They therefore highly value predictive information before reaching intersections. Within the I2Connect project, we developed a dynamic, evidence-based risk assessment for upcoming intersections. The system utilizes vehicle to everything (V2X) communication to collect information about the situation and predict the genral risk at the vehicle's arrival time at the intersection, displaying the risk indicator only when increased risk is predicted. The general risk is thereby defined through the common understanding that the more complicating factors are present at the same time, the more the chances of some unwanted incident happening increases. The individual risk increasing factors (e.g. visibility, road works, number of traffic participants, etc.) are thereby combined using Evidence Theory. A simulator study with professional truck drivers evaluated drivers general perception of such an indicator and compared two visualization alternatives; a dual-display shifting from dashboard to Head-Up Display (HUD) and a dashboard-only display. Drivers found the predictive alerts highly useful and preferred the HUD layout but also stated that such a system should adapt not only to the degree of risk but also to the driving situation as a whole (e.g. night time or how heavy the truck is loaded).

1. Introduction

The FFI-Vinnova provides these offer areas: "Support research and development activities that increase road safety and make transportation more sustainable through safe connected automated vehicles prepared for sharing." The project's contributions have aligned with FFI's overall impact goals, as it focuses on demonstrating innovative advanced driver assistance system (ADAS) solutions that substantial enhance road safety, driving efficiency, and environmental sustainability while ensuring that these benefits are accessible to all road users. And SAFER is this area as well.

Truck drivers have indicated that an ADAS function that supports intersection driving would be of great help. However, it is also known that ADAS functions are often perceived to be annoying and are switched off by roughly half of Europe's commercial drivers (Roberts, 2024).

The I2Connect project examins the integration of data derived from smart infrastructure and V2X communication through a theory-grounded data fusion subsystem. The aim is to enhance the acceptance of ADAS and significantly influence driving safety (Hesse et al., 2011).

In this paper, we present a situation aware risk assessment system. Thereby, the risk at the intersection is calculated for the vehicle's estimated time of arrival (ETA) and the risk indicator is only presented when an increased risk at ETA is predicted.

A multivariate simulator test with professional truck drivers was conducted. Firstly, our objective is to evaluate driver perception of the dynamic general risk indicator. Secondly, we want to understand user preference with regard to alternative HMI displays.

2. Method

2.1 Situation awareness through data fusion

To support the evaluation of different risk scenarios and to perform risk sensitivity analysis, we developed the risk assessment environment, SARA. It allows to display and analyse real-time or recorded data from a vehicle or any simulation environment.

SARA is based on a sensor-edge-cloud architecture (*Figure 1*) simulating the data pipeline of ADAS data flows. Smart Eye AB driver monitoring cameras provide sensor level gaze measurement patterns following the European New Car Assessment Programme's (Euro NCAP's) driver monitoring system (DMS) specifications. The external sensing layer comprises infrastructure-mounted cameras (Viscando, 2022) installed at urban intersections to track actor positions. This architeture divides tasks between device, cloud and edge nodes, which also can extract features locally to predict collisions risk (Atif et al., 2025).

A cooperative cloud platform aggregates evidence from multiple actor sources to achieve intersection-wide awareness. The system operates at 60Hz per data frame, which yields comprehensive snapshots of ego vehicle state, surrounding traffic, environmental conditions and driver attention data. The core element of SARA is the evidence-based risk assessment algorithm.

[1] https://www.his.se/en/research/informatics/interaction-lab/intelligent-driver-support-systems-and-safety-enhancement/

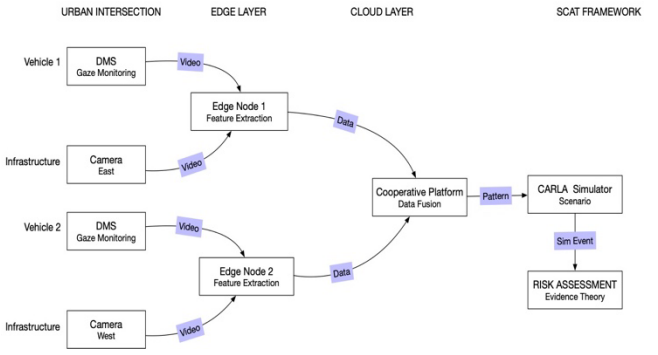


Fig. 1. *Overview of SARA's simulation framework*

Evidence theory (Shafer, 1976) is a generalization of probability theory (Dempster, 1969) that allows for fusion of uncertain data from multiple sources. It expresses the uncertainty regarding an unknown random variable, X, within a context of a given frame of discernment, Ω_X, through a probability mass function, m.

For risk prediction at the intersection, we consider number of traffic participants, presence of roadworks, number of drowsy or intoxicated drivers nearby, approaching emergency transports, and visibility. Each is described by a probability mass function. All mass functions are then combined, and the betting probability for the combined result is provided as the general risk. If it exceeds a preset threshold, the indicator is presented to the driver.

The SARA evaluation interface illustrating a live simulation run is shown in *Figure 2*. The left panel shows the risk development over time and its underlying probability mass distributions for the risk contributors. Below, SARA's parameter panel allows researchers to adapt fusion weights and thresholds on the fly without stopping the simulation. The upper panel presents indicators of vehicle dynamics. The right panel visualizes a running CARLA-based scenario, where the ego vehicle (red) navigates the intersection alongside vulnerable road users like pedestrians (white) and bicycles (green). In the top-right corner, the indicator status is highlighted and gaze classification (eyes-on-road vs. off-road) is overlaid, providing a real-time coupling between risk assessment output and the driver's attentional state. This dual panel visualizes the interdependencies between risk profile dynamics and driver behavior simultaneously, supporting systematic tuning of the investigated HMI concepts.



Fig. 2*. SARA's live dashboard*

2.2 HMI Concept Design

Drivers need different support at different stages while approaching an unregulated intersection. Therefore, we consider three distinct driving phases: the pre-planning phase (450 to 250 meters before the intersection), the planning phase (250 to 15 meters before the intersection), and the intersection phase (15 meters before to 15 meters after the intersection).

The layout of the digital cockpit strongly impacts information usability, the driver's mental workload, and driver safety (Smith et al., 2026). In our in-group comparative study, we conducted a multivariate test with two alternative HMI designs and a baseline with no HMI. Eighteen commercial heavy vehicle drivers were recruited. Three drivers were unable to complete the experiment, leaving 15 drivers who completed all three drives. Drives were randomly assigned and followed by a questionnaire after each drive, with a semi-structured interview at the end.

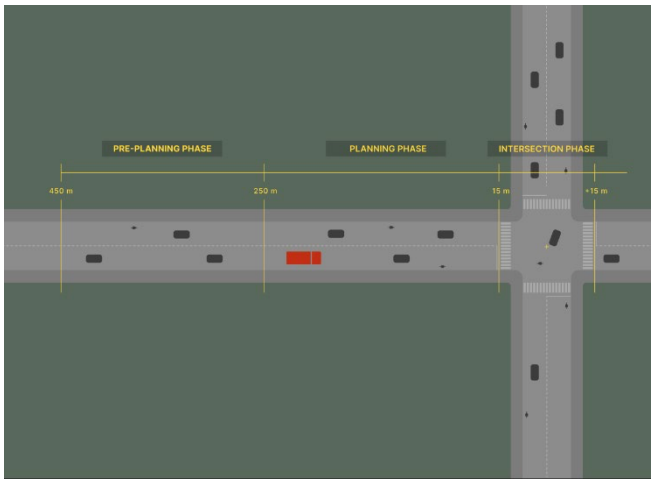


Fig. 3. *Intersection warning phases*

Concept A (Adaptive Dual-Display) showed the general risk indicator in the vehicle's dashboard while the vehicle was within the pre-planning phase and moved the indicator to the HUD when the vehicle entered the planning phase. The indicator remained in the HUD until the driver completely left the intersection.

Concept B (Simplified Single-Display) shows the general risk indicator in the vehicle's dashboard throughout all phases until the driver has left the intersection phase.

Baseline (no HMI) did not show any information about the predicted risk at the upcoming intersection.

3. Results

The majority of drivers perceived the risk indicator as helpful and, consistent with the findings from Liu et al (2023), preferred the indicator in the HUD. They described it to be easier to notice and to reduce the need to switch between looking at the dashboard and on the road, supporting them keeping their eyes on road, and providing a more natural and less distracting place for the information.

More than half of the drivers found the timing for the indicator appropriate and appreciated that it gave them enough time to prepare for the intersection. Some also mentioned that the disappearance of the indicator after the

intersection was suitable. Even if most drivers were positive to the timing within our experiment, the general consensus was that the timing should be adapted to the driving context.

Some drivers felt the indicator sometimes appeared too early, causing them to adapt their driving first and then readapt to the traffic situation, or forget about the information when they actually reached the intersection. One driver felt that an unverifiable risk indicator would make them feel unsafe. Some participants suggested that such an indicator was better suited for rural driving because in urban areas they already are highly alert to intersection risks.

While perceiving this particular information to be useful, many drivers were generally afraid of too much information and feared that the indicator could go off too often, even if this was not the case within our experiments.

Another appreciated concept was the combination of icon and text, as well as the coloring, where the yellow icon naturally alerted the driver without scaring them (as red might have done), and the text shortly explained the indication.

4. Conclusions

The qualitative nature of the experiment and the low number of participating drivers do not allow for statistically significant findings. At this stage, it cannot be concluded that the indicator influences traffic safety or driving behavior.

Furthermore, driver simulator experiments differ from real-world driving; what is perceived as helpful within the simulator might not be perceived that way in real traffic. At the same time, as Zhang et al. (2025) pointed out, while driving simulators are limited regarding physical feedback, they still have high relative validity for evaluating driver-HMI interaction and decision-making.

Our experiments showed that professional drivers of heavy vehicles perceive a general indicator for increased risk at upcoming intersections to be helpful. In agreement with Liu et al (2023), most drivers argued that the indicator is better suited within the HUD than in the dashboard.

Since drivers often state and literature confirms (e.g. Spahn et al. 2020) that there is no need for risk indicators if they already see the hazard, we must guard against over-warning, which can lead to alarm fatigue, system distrust, and eventual system deactivation. However, when a driver is distracted, an indication of even minor risks might be appropriate.

We conclude that a safety system for professional drivers, who often drive long hours and frequently experience fatigue, needs to be built catering to their specific needs and situations. To avoid annoying the driver, our next step will integrate driver monitoring. The system will suppress risk indicators when the driver is already attentive, while alerting even for minor risks if the driver is distracted. We must also balance minimizing information overload with transparency, keeping in mind that the evidence-based risk indicator means the driver might not always know why a risk is predicted.

Acknowledgments

The project basis is *Intelligent, Interactive and Connected (I2Connect) Next Generation of Realtime Driver Safety Support Systems* (2022-01652) FFI-Vinnova Traffic-safe automation, 2023-12-01 – 2026-05-3. Project partners: Univerisity of Skövde, Scania CV AB, Smart Eye AB, and Viscando, AB.

References

Atif, Yacine, Tarakanov, Yuri, Lebram, Mikael, Steinhauer, H. Joe, Karlsson Alexander, Hemeren, Paul. (2025). Data-Driven Prediction of Vehicle-Vulnerable Road User Collisions at Road Intersections Using Machine Learning Models. In Journal *Procedia Computer Science*, Volume 257, pp. 777-787. Elsevier.

Dempster, Arthur P. (1968). A Generalization of Bayesian Inference. Journal of the Royal Statistical Society: Series B (Methodological) 30, 2 (1968), 205–232

Hesse, T., Engström, J., Johansson, E., Varalda, G., Brockmann, M., Rambaldini, A., ... & Kanstrup, L. (2011). Towards user-centred development ofintegrated information, warning, and interventionstrategies for multiple ADAS in the EUproject interactive. Springer Berlin Heidelberg.

Liu, Tong , Shen, Jiahui, Xue, Gang, Zhang, Jiahao, Ma, Keli and Zhang, Wei. (2023). Does the AR-HUD system affect driving behaviour? An eye-tracking experiment study. Transportation Research Interdisciplinary Perspectives 20 (2023), 101150. doi:INTERNALSTYLE-FILEERROR

Roberts, G. (2024). One in Two Drivers Switch Off Driver Assistance Systems. Fleet News. https://fleetnews.co.uk/drivers-switch-off-driver-assistance-systems

Shafer, G. (1976). A Mathematical Theory of Evidence., Princeton University Press

Smith, J. L., Gabbard, J. L., & Conley, C. (2016). Head-up vs. head-down displays: Examining traditional methods of display assessment while driving. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive'UI '16)*, 185–192.

Spath, Dieter, Weber, Betina, Dangelmaier, Manfred and Diederichs, Frederik. (2020). User rating and acceptance of attention-adaptive driver safety systems. European Transport Research Review 12, 1 (2020), 1–11. doi:10.1186/s12544-020-00414-w

Viscando AB. (2022). OTUS3D: 3D and AI-based infrastructure sensor for traffic monitoring and road user behaviour analysis. *Viscando AB. Retrieved March 13, 2026*

Zhang, Siyang, Zhao, Chi, Zhang, Zherui and Lv, Yecheng (2025). Driving simulator validation studies: A systematic review. Simulation Modelling Practice and Theory 138 (2025), 103020. doi:10.1016/j.simpat.2024.103020

Drowsiness and sleep prediction via imaging PPG: a multiscenario validation from Simulated to Naturalistic Driving

Michele Guagnano [1], Sara Groppo [1*], Andrea Romigi[2], Riccardo Groppo[3], Massimo Violante [1]

[1] Politecnico di Torino, Department of Control and Computer Engineering (DAUIN), Corso Castelfidardo, 34/D, 10138, Torino (TO), Italy
[2] Saint Camillus International University of Health and Medical Sciences, Via di Sant'Alessandro 8, 00131, Roma (RM), Italy
[3] Sleep Advice Technologies srl, Corso Vinzaglio, 12, 10121, Torino (TO), Italy

[*] sara.groppo@polito.it

Abstract: Driver drowsiness is a major contributor to road accidents, accounting for an estimated 10–30% of all vehicle crashes. Driver Monitoring Systems (DMS) aim to assess driver alertness and are typically categorized into vehicle-based, behavioral, and physiological approaches. Among these, physiological monitoring provides the most reliable indicators of drowsiness; however, many existing systems rely on contact-based sensors that are intrusive and impractical for real-world use. This study investigates a fully contactless physiological monitoring approach based on imaging photoplethysmography (iPPG) to detect sleep-related events in drivers. The proposed method extracts cardiovascular signals from video recordings, enabling unobtrusive monitoring of the driver's physiological state in both controlled and naturalistic driving conditions. Data were collected across multiple experimental scenarios. Six participants underwent a 45-minute Maintenance of Wakefulness Test (MWT), four participants completed a 45-minute static driving simulation, and four participants performed a 45-minute dynamic driving simulation. Sleep-related events were objectively validated using polysomnography (PSG) recordings reviewed by a medical expert according to the American Academy of Sleep Medicine (AASM) guidelines. Additionally, five participants performed real-world driving sessions, where sleep events were validated using a previously developed smartwatch-based sleep-detection algorithm. The heuristic algorithm achieved an overall accuracy of 88.6%, with 82.4% specificity (95% CI: 56.6%–96.2%) and 92.6% sensitivity (95% CI: 75.7%–99.1%), demonstrating the feasibility of a non-intrusive physiological solution for driver sleep event detection.

1. Introduction

Globally, it is estimated that road accidents result in approximately 1.35 million deaths and up to 50 million injuries annually (Shaik et al., 2021). Drowsy driving is a primary risk factor for accidents, contributing approximately to 10–30% of all road traffic incidents and a significant cause of fatal accidents (Gonçalves et al., 2015). Drowsiness-related accidents are typically more severe due to their occurrence on high-speed roads. Sleep deprivation and prolonged driving are key factors. Given its serious consequences, early detection of drowsiness through technology could significantly reduce traffic-related deaths and injuries (Ebrahim Shaik, 2023).

The Driver Monitoring Systems (DMS) play a crucial role in detecting the state of the driver and can be categorized as behavioral, vehicular, and physiological-based techniques (Sahayadhas et al., 2012). Among them, the analysis of physiological signals is the most reliable method, but it can be intrusive if contact-based, sensitive to noise and motion, and costly (Perkins et al., 2023; Ramzan et al., 2019).

In this context, imaging photoplethysmography (iPPG) represents a promising contactless alternative for driver monitoring. iPPG allows the extraction of cardiovascular signals from subtle variations in skin color captured by standard cameras, enabling the estimation of HR and HRV without physical contact (McDuff et al., 2014; Verkruysse et al., 2008). Thanks to its unobtrusive nature, this technology can be easily integrated into in-vehicle camera systems and continuously monitor the driver's physiological state. These characteristics make iPPG particularly suitable for real-world driver monitoring applications, where comfort, usability, and long-term acceptance are essential.

In-vehicle drowsiness validation is typically validated by exploiting the driver's subjective self-assessments (eur-lex.europa.eu/legal-content). However, it was demonstrated that subjective measures often failed to accurately reflect actual drowsiness levels, capturing only the driver's perceived state (Anderson et al., 2023; Cai et al., 2021; Groppo et al., 2024).

This study investigates the prediction of drowsiness, microsleep, and sleep events in both simulated and naturalistic driving scenarios. The proposed heuristic algorithm exploits HR extracted through iPPG. Objective ground truth was established using polysomnography (PSG) recordings reviewed by a medical expert according to the American Academy of Sleep Medicine (AASM) guidelines, as a validation method strongly assessed in a previous study (Groppo et al., 2026).

Method

1.1 Participants and experimental protocol

19 participants were enrolled (6 females; mean age 33 ± 13 years; mean BMI 24 ± 3 kg/m2). Data were collected across multiple experimental scenarios designed to evaluate the proposed approach under both controlled and real-world conditions.

The following tests were done under controlled conditions:

- six participants underwent a Maintenance of Wakefulness Test (MWT), remaining seated on a comfortable chair in a quiet room not moving for 45 minutes;
- four participants completed a 45-minute static driving simulation;
- four participants performed a 45-minute dynamic driving simulation.

To obtain an objective ground truth for sleep-related events during the controlled experiments, every participant was equipped with a full PSG setup (Noxturnal A1). The PSG recording was subsequently reviewed by a medical expert according to the guidelines of the AASM.

To further assess the feasibility of the proposed approach in real-world conditions, five participants also performed real driving sessions. In this case, considering that PSG would have been too invasive, sleep events were identified using a sleep-detection algorithm previously developed and validated by the same authors, running in real time on a smartwatch.

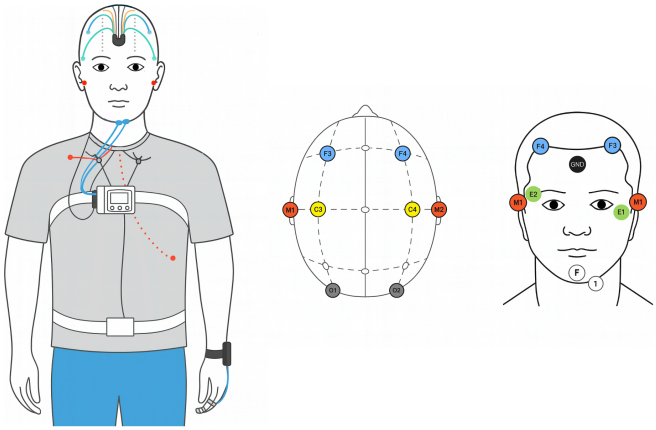


Figure 1 - PSG setup for sleep validation in drivers, comprising EEG (GND, F4, F3, C3, C4, O1, O2, M1, M2), EOG (E2, E1), EMG (F, I), ECG, two respiratory belts, and a PPG sensor.



Figure 2 - *Experimental setups: MWT, static driving simulator, dynamic driving simulator, and real-vehicle configuration. The iPPG device was positioned in front of the participant on the right side.*

1.2 HR based sleep prediction algorithm

The process begins with the remote extraction of the PPG signal from the subject's facial regions using a near-infrared (NIR) LGE camera. From this optical signal, two primary physiological inputs are derived: Heart Rate (HR) and Beat-to-Beat Intervals (BBI). The algorithm's operation is shown in the following pseudocode.

```
START
// Data Acquisition
while there is signal do
  Collect samples of HR and BBI
  if n samples are collected then
    Calculate the average of HR and the std of BBI in the window
    Store them in mHR and sBBI
    return Level0: Data acquisition;
// Data Processing
    if N values are calculated then
      The windows slide in time discarding oldest value
      and adding new value
      Calculate the first derivative of sBBI and store in dsBBI
      count = 0
      DOD = 0
      for i = 1 to i = WindowSize-1 do
        if dsBBI[i] < sTh and mHR < mTh then
          count++
        else
          count = 0
        end if
        if count > DOD then
          DOD = count
        end if
      end for
// Classification
      if DOD <= Th1 then
        return Level1: Awake;
      end if
      if DOD <= Th2 and DOD > Th1 then
        return Level2: Fatigue Onset;
      end if
      if DOD > Th2 then
        return Level3: Drowsiness Onset;
      end if
    end if
  end if
end while
```

Figure 3 – *Pseudocode of the algorithm: Data acquisition, Data processing, and Classification stages summarized.*

2. Results

Building on the multi-scenario experimental framework, the algorithm's performance was evaluated across a three-level classification scale: Wakefulness (Level 1), Fatigue (Level 2), and Sleep/Microsleep events (Level 3). The system was designed to generate two distinct alerts: a Fatigue Onset trigger, aimed at predicting the transition to intermediate fatigue, and a Drowsiness Onset trigger, specifically calibrated to predict imminent sleep or microsleep episodes.

Validation against PSG ground truth and reference sensors demonstrated high diagnostic accuracy. As shown in Table 1, the algorithm correctly identified 25 critical events (True Positives), including both fatigue and sleep transitions, and successfully classified 14 sessions of sustained wakefulness (True Negatives). Despite the increased complexity of dynamic simulation and on-road trials, the prediction of both onset types remained robust. Out of the nine dynamic/real-world sessions where sleep was expected, the Fatigue Onset was correctly triggered in 100% of detected events, with only two False Negatives across the entire 19-subject dataset (Mwt_P2 and Dynamic_P3). Statistical evaluation yields a global Sensitivity of 92.6% and a Specificity of 82.4%, with an overall Accuracy of 88.6%.

3. Discussion

The discrepancy in Static_P3 (3 False Positives) stems from the participant's use of beta-blockers, which pharmacologically reduced Heart Rate Variability (HRV). This "flattened" cardiac profile misled the algorithm into detecting a fatigue state despite active wakefulness. The low False Positive rate is particularly significant for real-world applications. By minimizing alarms, the proposed system preserves driver trust and avoids the 'alarm fatigue' that often leads users to disregard or disable safety monitoring technologies.

Regarding the two missed detections (Mwt_P2 and Dynamic_P3), it is critical to note that both tests featured darker skin tones. In iPPG applications, higher melanin levels can attenuate reflected light, potentially reducing the signal-to-noise ratio (SNR) during abrupt transitions. Furthermore, these specific events were characterized by sudden microsleeps. While they did not trigger a Drowsiness Onset (Level 3) alert, they were both consistently preceded by a Fatigue Onset (Level 2) notification. This suggests that while signal quality or the abruptness of the event prevented the system from reaching the stricter Level 3 threshold, the algorithm successfully identified the preceding physiological degradation.

From a safety perspective, this hierarchical alerting system ensures that even when a microsleep is not explicitly labeled, the driver has already been warned of a high-fatigue state. This confirms the algorithm's effectiveness as a multi-stage early warning system, capable of tracking the physiological drift toward impairment even under challenging optical, pharmacological, and dynamic conditions.

4. Conclusions

Drowsy driving is a primary risk factor for accidents, contributing approximately to 10–30% of all road traffic incidents. The analysis of physiological signals is the most reliable method for Driver Monitoring Systems (DMS), but it can be intrusive, sensitive to noise and motion, and costly. To date, validation methods for driver impairments are based on subjective self-assessments, which often fail.

This study validates a contactless, near-infrared iPPG-based framework for monitoring driver impairment across a spectrum of physiological states. By implementing a hierarchical classification of Fatigue and Drowsiness Onsets, the system offers a robust multi-stage early warning mechanism that remains effective even under challenging dynamic and real-world conditions. While pharmacological factors (beta-blockers) and physiological traits (darker skin tones) may influence specific detection thresholds, the algorithm consistently identifies the preceding degradation in alertness. These results demonstrate that remote cardiac imaging is a viable and reliable alternative to wearable sensors, providing a critical safety layer for the prevention of sleep-related accidents in automotive applications.

5. Acknowledgments

The authors gratefully acknowledge LG Electronics (LGE) for providing the near-infrared iPPG camera system used for data acquisition in this study.

References

Anderson, C., Cai, A. W. T., Lee, M. L., Horrey, W. J., Liang, Y., O'Brien, C. S., Czeisler, C. A., & Howard, M. E. (2023). Feeling sleepy? stop driving—awareness of fall asleep crashes. *Sleep*, *46*(11). https://doi.org/10.1093/sleep/zsad136

Anniss, A. M., Young, A., & O'Driscoll, D. M. (2021). Microsleep assessment enhances interpretation of the Maintenance of Wakefulness Test. *Journal of Clinical Sleep Medicine*, *17*(8), 1571–1578. https://doi.org/10.5664/jcsm.9250

Cai, A. W. T., Manousakis, J. E., Lo, T. Y. T., Horne, J. A., Howard, M. E., & Anderson, C. (2021). I think I'm sleepy, therefore I am – Awareness of sleepiness while driving: A systematic review. In *Sleep Medicine Reviews* (Vol. 60). W.B. Saunders Ltd. https://doi.org/10.1016/j.smrv.2021.101533

Doghramji, K., Mitler, M. M., Sangal, R. B., Shapiro, C., Taylor, S., Walsleben, J., Belisle, C., Erman, M. K., Hayduk, R., Hosn, R., O'malley, E. B., Sangal, J. M., Schutte, S. L., & Youakim, J. M. (1997). *A normative study of the maintenance of wakefulness test (MWT)*.

Ebrahim Shaik, M. (2023). A systematic review on detection and prediction of driver drowsiness. In *Transportation Research Interdisciplinary Perspectives* (Vol. 21). Elsevier Ltd. https://doi.org/10.1016/j.trip.2023.100864

Gonçalves, M., Amici, R., Lucas, R., Åkerstedt, T., Cirignotta, F., Horne, J., Léger, D., Mcnicholas, W. T., Partinen, M., Téran-Santos, J., Peigneux, P., & Grote, L. (2015). Sleepiness at the wheel across Europe: A survey of 19 countries. *Journal of Sleep Research*, *24*(3), 242–253. https://doi.org/10.1111/jsr.12267

Groppo, S., Guagnano, M., Pugliese, L., Sini, J., & Violante, M. (2024). To be or not to be... awake? A comparison of subjective and objective methods for drowsiness detection in drivers. *2024 IEEE 22nd Mediterranean Electrotechnical Conference, MELECON 2024*, 684–

688. https://doi.org/10.1109/MELECON56669.2024.10608467

Groppo, S., Guagnano, M., Serchi, V., Beltrao, G., Pugliese, L., & Violante, M. (2026). Contactless System for Sleep Prediction in Drivers. *IEEE Access*. https://doi.org/10.1109/ACCESS.2026.3667411

Hertig-Godeschalk, A., Skorucak, J., Malafeev, A., Achermann, P., Mathis, J., Schreier, D. R., & David Schreier, med. (2019). *Microsleep episodes in the borderland between wakefulness and sleep*. https://doi.org/10.1093/sleep/zsz163/5536744

Lu, K., Sjörs Dahlman, A., Karlsson, J., & Candefjord, S. (2022). Detecting driver fatigue using heart rate variability: A systematic review. *Accident Analysis and Prevention*, *178*. https://doi.org/10.1016/j.aap.2022.106830

McDuff, D., Gontarek, S., & Picard, R. W. (2014). Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Transactions on Biomedical Engineering*, *61*(10), 2593–2601. https://doi.org/10.1109/TBME.2014.2323695

Perkins, E., Sitaula, C., Burke, M., & Marzbanrad, F. (2023). Challenges of Driver Drowsiness Prediction: The Remaining Steps to Implementation. *IEEE Transactions on Intelligent Vehicles*, *8*(2), 1319–1338. https://doi.org/10.1109/TIV.2022.3224690

Ramzan, M., Khan, H. U., Awan, S. M., Ismail, A., Ilyas, M., & Mahmood, A. (2019). A Survey on State-of-the-Art Drowsiness Detection Techniques. In *IEEE Access* (Vol. 7, pp. 61904–61919). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2019.2914373

Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: A review. In *Sensors (Switzerland)* (Vol. 12, Number 12, pp. 16937–16953). MDPI AG. https://doi.org/10.3390/s121216937

Shaik, M. E., Islam, M. M., & Hossain, Q. S. (2021). A review on neural network techniques for the prediction of road traffic accident severity. In *Asian Transport Studies* (Vol. 7). Elsevier Ltd. https://doi.org/10.1016/j.eastsj.2021.100040

Tondo, P., Pronzato, C., Risi, I., Perretti, C., De Gennaro, L., Bonsignore, M. R., Malovini, A., & Fanfulla, F. (2024). *Downloaded from jcsm.aasm.org by Francesco Fanfulla on*.

Verkruysse, W., Svaasand, L. O., & Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. *Optics Express*, *16*(26), 21434. https://doi.org/10.1364/oe.16.021434

Appendix A

Table 1 Prediction results

Test_ID	TP	FP	TN	FN	Notes
Static_P1	0	0	1	0	
Static_P2	0	0	1	0	
Static_P3	2 (2F)	3	0	0	Takes Betablockers
Static_P4	0	0	1	0	
Static_P5	1 (1F)	0	0	0	Dark skin
Static_P6	0	0	1	0	
Mwt_P1	5 (5F)	0	0	0	
Mwt_P2	11 (4M, 1S, 6F)	0	0	1 (1M)	Dark skin
Mwt_P3	0	0	1	0	
Mwt_P4	0	0	1	0	
Dynamic_P1	0	0	1	0	
Dynamic_P2	0	0	1	0	
Dynamic_P3	6 (6F)	0	0	1 (1M)	Dark skin
Dynamic_P4	0	0	1	0	
Real_P1	0	0	1	0	
Real_P2	0	0	1	0	
Real_P3	0	0	1	0	Takes Betablockers
Real_P4	0	0	1	0	
Real_P5	0	0	1	0	Dark skin
Total	25	3	14	2	

Abbreviations: TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative
F = Fatigue, M = Microsleep, S = Sleep Event

A Methodological Approach to Evaluate Driver Distraction During In-vehicle Interactions with Voice-based Large Language Models

Nadja Schömig[1*], Christina Kremer[1], Thomas Brand[1], Denise Sogemeier[2], Frederik Naujoks[2], Alexandra Neukum[1], Andreas Keinath[2]

[1] Wuerzburg Institute for Traffic Sciences (WIVW GmbH), Robert-Bosch-Strasse 4, Veitshöchheim, Germany
[2]BMW Group, Knorrstraße 147, Munich, Germany
[*]nadja.schoemig@wivw.de

Abstract: This study presents a methodological approach for evaluating driver distraction during voice-based in-vehicle Large Language Model (LLM) interactions. We adapted established test protocols for visual-manual distraction to accommodate voice-based interactions while assessing both visual and cognitive distractions. In a driving simulator, we examined natural conversations with the LLM which differed in feedback modality (voice only vs. voice + visual) and feedback length (short vs. long). The results showed that all use cases met the absolute glance criteria in terms of individual glance duration, while the total eyes-off-road time criterion was only met for the condition "voice only, long texts". Driving performance and cognitive distraction (measured by tactile DRT) were comparable to the reference task "hands-free phone call". Additional analyses suggested that speech production was particularly cognitively demanding, but nevertheless remained within acceptable limits. The study demonstrates that the proposed methodological approach is reasonable for evaluating visual and cognitive distraction effects of LLM interactions, except for the criterion of total eyes-off-road time, which proved unsuitable for longer interactions.

1. Introduction

The rapid integration of Large Language Models (LLMs) into automotive environments has introduced novel challenges and risks associated with driver distraction. Although LLMs offer substantial potential for enhancing information access and user experience, the visual and cognitive demands of voice-based LLM interactions remain poorly understood. While established test methods exist for evaluating distraction by visual-manual tasks (e.g., AAM, 2006; NHTSA, 2013), comparable methodologies with acceptance criteria for voice systems remain lacking. Although voice systems have shown to be less visually demanding than visual-manual distraction sources (e.g., Monk et al., 2023), the role of cognitive distraction should be emphasized, as both speech production and auditory processing of system responses can impose significant cognitive load. Understanding cognitive distraction is crucial for assessing safety implications of LLM-based interfaces.

This study presents a methodological approach for evaluating driver distraction during in-vehicle LLM interactions which aims to evaluate not only the impact on visual distraction using established criteria, but also the influence on cognitive distraction and, consequently, on driving performance. The approach was adapted from existing test protocols originally developed for visual-manual tasks (AAM, 2006; NHTSA, 2013) to accommodate voice-based LLM interactions. Additionally, the study examined whether feedback modality and feedback length influence visual and cognitive driver distraction.

2. Method

2.1 LLM use cases

The study examined use cases (UC) in which drivers verbally interacted with a LLM system (ChatGPT) while driving. Participants were asked to query the system on four standardized topics (e.g. tips for a holiday trip to a warm destination, or ideas for a birthday party) and to maintain the respective conversation for a duration of three minutes. The spoken utterances were visually displayed on the center information display (CID). The current system status and, depending on the test conditions, a visual transcript of the voice output were also displayed here. Four LLM UCs were defined by varying the factors “feedback modality” (voice only vs. voice + visual) and “length of feedback” (short, 120-140 characters vs. long, 400-420 characters).

2.2 Test setting

The study was conducted in the static WIVW driving simulator (Figure 1).



***Fig. 1**: Implementation of the LLM in the WIVW driving simulator.*

The driving task was a continuous car-following task (adapted from NHTSA, 2013, 2019) with participants following a lead vehicle at a speed of 100 km/h with a constant distance of 50 m while maintaining good lane control. Cognitive load was assessed using the tactile DRT in accordance with ISO 17488 (ISO, 2016). Participants were asked to respond as quickly and accurately as possible to vibrations from a motor attached to their shoulder by pressing a microswitch (Figure 2).

The driving task and the DRT were practiced individually and in combination with each other sufficiently before the test drive.

The test drive (approx. 25 minutes) consisted of a total of 6 sections, each with a fixed duration of 3 minutes: 4 sections with LLM UCs, a baseline section (consisting of car following and detection response task (DRT)), and a reference section (hands-free phone call with the experimenter).

2.3 Dependent measures and data analysis

Glance data were analyzed with respect to the fulfilment of the absolute criteria defined in NHTSA (2019): For at least 21 of the 24 test participants, the following criteria had to be met:

A. no more than 15% of off-road glances exceed 2.0 s
B. the mean off-road glance duration is ≤ 2.0 s
C. the total eyes-off-road time (TEORT) is ≤ 12.0 s.

To investigate the impact of distraction caused by interaction with the LLM on driving performance, we used dependent *t*-tests with Bonferroni-Holm adjusted *p*-values to compare the standard deviation of lateral position (SDLP) and standard deviation of car follow distance (SD distance) with those from the reference task. The criterion for task acceptance was that the performance in the LLM UCs should be equal or better compared to the reference task.



***Fig. 2**: Vibrotactile transponder and microswitch for the tactile DRT.*

The same approach was applied to the dependent measures DRT response time and miss rate.

To examine the influence of the factors feedback modality and feedback length on the reported measures, repeated-measures ANOVAs were also performed.

2.4 Study sample

The study sample consisted of 24 participants (n=12 female), evenly distributed to four age groups (NHTSA, 2013). Mean age was 40.9 years (SD=14.7). All drivers were trained in the simulator. A prerequisite for participation in the study was previous experience with voice assistants.

3. Results

3.1 Fulfillment of distraction evaluation criteria

The analysis of eye glance data showed that glance criteria A and B were met for all UCs. Criterion C (TEORT) was fulfilled only for UC "voice only, long texts".

Lateral and longitudinal driving performance SDLP and SD Distance for all four LLM UCs was at the same level as for the reference task. Also, DRT performance for all four LLM UCs was on an equal level or better compared to the reference task (Table1. Figure 3, Figure 4).

3.2 Effect of feedback modality and feedback length

The results from the repeated-measures ANOVAS revealed a significant main effect of feedback modality on mean glance duration (Table 2). The UCs with voice and visual feedback resulted in longer glances than the voice

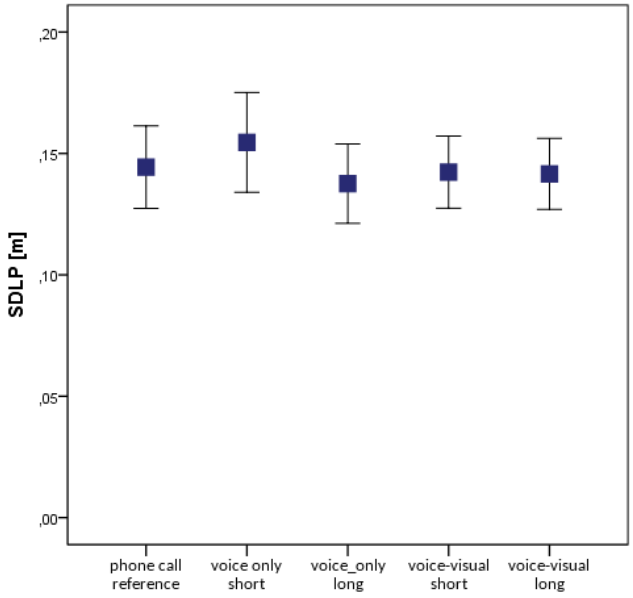
***Table 1:** Results of dependent t-tests between each LLM UC with reference task.*

	Driving performance				**DRT**			
	SDLP		**SD Distance**		**Response time**		**Miss rate**	
Reference task vs.	***t*** (23)	***p***	***t*** (23)	***p***	***t*** (23)	***p***	***t*** (23)	***p***
Voice only; short	1.216	0.944	-1.280	.684	0.262	>.999	-1.480	.266
Voice only; long	-0.937	>.999	-1.412	.684	-1.581	.384	-3.317	.012*
Voice-visual; short	-0.395	>.999	-0.632	.818	-0.349	>.999	-1.558	.266
Voice-visual; long	-.0412	>.999	-0.841	.818	-2.126	.176	-2.842	.030*

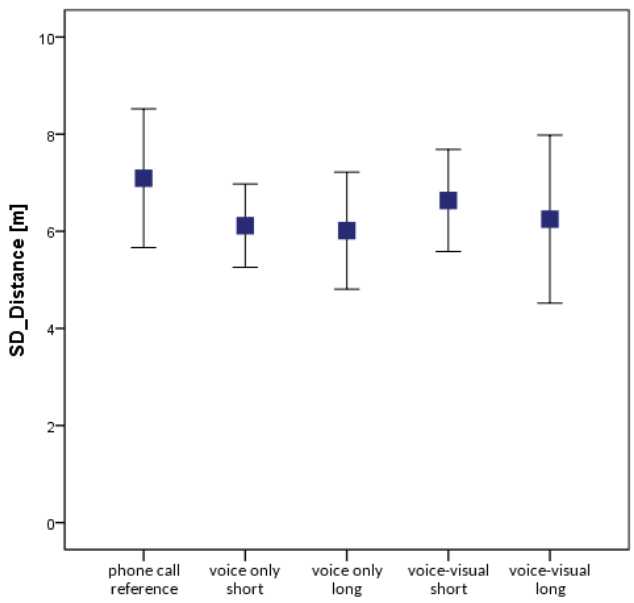
Note. p-values were adjusted using the Bonferroni-Holm correction.

Table 2: *Results of repeated-measures ANOVAS for "feedback modality" and "feedback length".*

	Glance measures				Driving performance				DRT performance			
	Mean glance duration		Number of glances		SDLP		SD Distance		Response time		Miss rate	
Factors	***F*** (1,21)	***p***	***F*** (1,21)	***p***	***F*** (1,23)	***p***	***F*** (1,23)	***p***	***F*** (1,23)	***p***	***F*** (1,23)	***p***
Feedback modality	24.5	.001 ***	4.23	.052	.388	.539	.749	.396	1.05	.315	.122	.730
Feedback length	1.98	.174	1.56	.225	1.91	.180	.291	.595	7.39	.012*	3.16	.089
Modality x length	0.22	.646	.003	.875	.310	.096	.080	.780	.028	.870	.280	.601

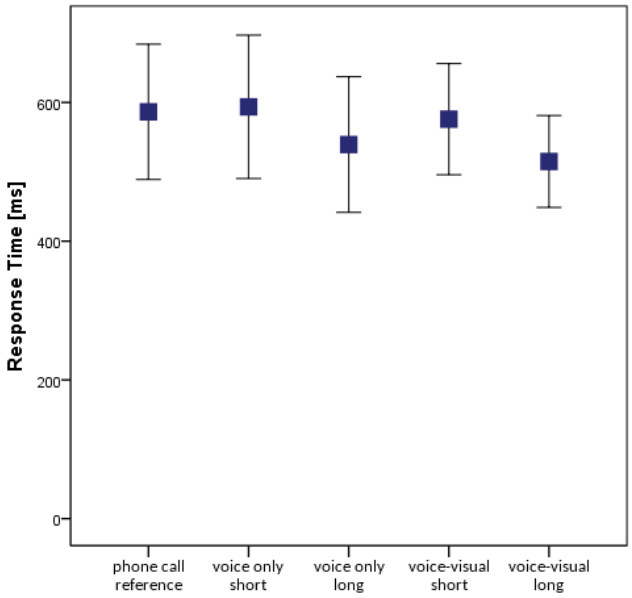


a

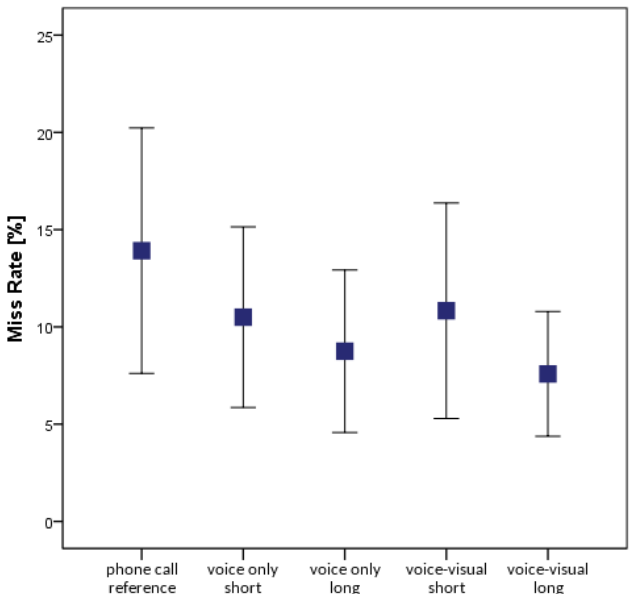


b

Fig. 3*: Driving performance data for LLM UCs and the reference task. (a) SDLP (Mean ± 95% CI), (b) SD Distance (Mean ± 95% CI)*



a



b

Fig. 4*: DRT performance data for LLM UCs and the reference task.(a) Response time (Mean ± 95% CI), (b) Miss rate (Mean ± 95% CI)*

only condition. There was a marginally significant effect of feedback modality on the number of glances with tendentially more glances to the CID with additional visual feedback. Thus, although the system's responses were not visually displayed in the "voice only" condition, the spoken utterances resulted in multiple glances at the CID.

No influences of feedback modality and feedback length on driving performance measures were found. This also applies to the miss rate in DRT. There was a significant effect of feedback length on DRT response times: Shorter system outputs led to longer response times (resulting from a higher frequency of interaction with the system).

3.3 Distribution of misses in the DRT

A detailed analysis of DRT miss rates across LLM interaction phases revealed that speech production phase (20-25% of total UC duration, from button press to system feedback) accounted for 40-45% of all misses. The overrepresentation of DRT misses in this phase indicates that speech production during LLM interaction is particularly cognitively demanding.

4. Discussion

Taken together, all examined LLM UCs met the defined distraction evaluation criteria except TEORT. This is reasonable as the UC duration was 3 minutes whereas the TEORT threshold of 12 s was originally defined for visual-manual tasks with clearly shorter durations (e.g., manual radio tuning). Therefore, all investigated LLM UCs appear to be safe in automotive contexts.

Displaying text in addition to voice output increases glance duration and number of glances, but not to a critical extent (individual glance duration still meets acceptance criteria). Different feedback lengths showed similarly low visual distraction effects indicating that drivers did not actually read the text. Notably, drivers also glanced at the CID during purely verbal output to verify spoken utterances visually. The statistical analysis indicated that shorter LLM responses actually tend to increase distraction —likely due to higher interaction frequency within the constant interaction duration, leading to increased speech production and visual control needs.

5. Conclusions

The presented methodological approach reasonably evaluates distraction effects of in-vehicle LLM interactions by assessing both visual and cognitive distraction effects through glance metrics and tactile DRT performance. Effects on driving performance enable statements about potential safety implications.

For long LLM interactions, the TEORT criterion appears unsuitable and should not be used for distraction evaluation.

6. Acknowledgments

This research has been funded by the BMW Group. All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations. We thank the editors and reviewers for their valuable time and effort.

References

Alliance of Automobile Manufacturers (2006). Statement of Principles, Criteria, and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems. Washington, D.C.

International Organization for Standardization (2016). Road vehicles - Transport information and control systems – Detection-response task (DRT) for assessing attentional effects of cognitive load in driving (ISO Standard No. 17488). Retrieved from https://www.iso.org/standard/59887.html

Monk, C., Sall, R., Lester, B. D., & Higgins, J. S. (2023). Visual and cognitive demands of manual and voice-based driving mode implementations on smartphones. Accident Analysis & Prevention, 187, 107033.

National Highway Traffic Safety Administration (2013). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).

NHTSA (2019). Visual-Manual Driver Distraction Guidelines. Test Procedures – Task Acceptance Testing. Occlusion and Eye Glance Measurement Using Driving Simulator Testing (DOT HS 812 739).

Appendix

Table 1: *Mean (M) and SD for all measures and all UC.*

Eye glance metrics	**Mean glance duration [ms]**		**Mean number of glances [n]**	
	M	**SD**	**M**	**SD**
Voice only; short	641	537	13.0	9.47
Voice only; long	645	433	11.78	11.15
Voice-visual; short	817	505	17.33	12.89
Voice-visual; long	953	605	16.48	12.72

Eye glance metrics	**Total Eyes-off road time [s]**	
	M	**SD**
Voice only; short	8.33	7.31
Voice only; long	7.60	8.14
Voice-visual; short	14.11	12.15
Voice-visual; long	15.71	12.62

Driving performance	**SDLP [m]**		**SD Distance [m]**	
	M	**SD**	**M**	**SD**
Baseline	.161	.087	4.29	1.57
Reference task	.144	.040	7.09	3.39
Voice only; short	.155	.049	6.12	2.03
Voice only; long	.138	.039	6.01	2.85
Voice-visual; short	.142	.035	6.63	2.49
Voice-visual; long	.142	.035	6.25	4.09

DRT performance	**Response Time [ms]**		**Miss Rate [%]**	
	M	**SD**	**M**	**SD**
Baseline	323	83	.042	1.53
Reference task	586	231	13.92	14.94
Voice only; short	594	245	10.50	10.99
Voice only; long	539	231	8.75	9.89
Voice-visual; short	576	190	10.83	13.12
Voice-visual; long	515	157	7.58	7.59

Hands On, Eyes On! Driver Monitoring and Behaviors in Hands-Free Level 2 Automation

Linda Pipkorn[1*], Pnina Gershon[2], Bruce Mehler[2], Bryan Reimer[2],

[1] Chalmers University of Technology, Mechanical Engineering, Hörsalsvägen 5, Gothenburg, Sweden
[2] Massachusetts Institute of Technology, MIT AgeLab, 1 Amherst Street, Cambridge, USA
[*]linda.pipkorn@chalmers.se

Hands-free Level 2 (L2) systems are now widely deployed in the US. Controlled studies suggest that reduced manual control may induce driver overreliance; however, real-world usage remains understudied. This study examines driver behavior when using Ford Mach-E's hands-free L2 system, BlueCruise. Both hands-on-wheel and eyes-on-road alerts occurred, with hands-on-wheel alerts more frequent (about two per hour), while eyes-on-road alerts were more strongly associated with individual characteristics. Alerts escalating to a request to resume control were rare (0.03%). For both alert types, alerts were not consistently triggered during non-compliance (eyes off road or hands off wheel). When non-complient, drivers typically looked back to the road quickly (0.5 s) but returned their hands more slowly (2.1 s). More hands-off driving was also associated with more eyes-on-road alerts and longer glances to the down and center-stack regions. Future research should examine this tendency to better understand the benefits and risks of hands-free driving.

1. Introduction

Commercial hands-free L2 driving automation systems are now widely available to consumers in USA (Ford, 2025). These systems provide lateral and longitudinal control within defined operational design domains (ODD; SAE International, 2018) but require drivers to remain engaged and ready to intervene. Although automation can improve comfort and reduce workload, it may also increase driver overreliance and out-of-the-loop performance, particularly as continuous manual control is reduced (Merat et al., 2019; Victor et al., 2018).

Drivers rapidly adopt L2 automation when it becomes available (Gershon et al., 2021), and having hands off the wheel has been shown to delay responses to system-initiated disengagements (Gershon et al., 2023). At the same time, hands-free systems shift the primary engagement requirement from continuous manual control to visual monitoring, placing greater emphasis on driver monitoring systems and alerting strategies. However, real-world data on how drivers interact with these alerts during early use remain limited.

This study examines drivers' real-world use of the hands-free L2 system BlueCruise in the Ford Mustang Mach-E during a four-week exposure. Focusing on glance and steering behaviour in relation to driver monitoring alerts, the study quantifies alert frequency, response times, and associations between hands-off and eyes-off behaviour. Understanding these interactions is important for informing the design and calibration of driver monitoring and alerting strategies in hands-free L2 systems.

The following research questions are considered:

- **RQ1:** What is the prevalence (number of alerts/hours of highway driving) of hands-on-wheel and eyes-on-road alerts?
- **RQ2:** How do drivers respond to alerts (visual and physical response times)?
- **RQ3:** Is there a correlation between driving more hands-off and driving more eyes-off?

1. Method

1.1 Participants and dataset

Data were sampled from the ongoing MIT Advanced Vehicle Technology project in the New England region of the US. Eleven drivers (3 female), mean age 45yrs (SD=8), were included in this analysis. The data constitutes 167 hours of highway driving.

1.2 Data processing, annotation and analysis

Annotation of glance and steering wheel control level followed standard MIT AgeLab annotation procedures (Morando et al., 2021; Pipkorn et al., 2025), while the annotation of driving mode and alert presence used instrument cluster display video (Fig. 1) (see Tab. 1).



Fig. 1 *The instrument panel display in the Ford Mach-E, displaying BlueCruise hands-free mode.*

Tab. 1. *Instrument cluster annotations*

Driving mode	manual, BlueCruise hands-free, BlueCruise hands-on or Adaptive Cruise Control (ACC).
Hands-on-wheel alerts	(i) first-stage alert (alert in display with orange frame), (ii) second-stage alert (alert in display with red frame and audio tone). Both alerts included the message "Keep hands on steering wheel".
Eyes-on alerts	(i) first-stage alert (alert in display with white frame or alert in display with orange frame and audio tone reading). Both alerts included the message "Watch the road"
Resume-control alerts	alert in display with red frame reading "resume control".

Each highway trip was used to extract: (a) presence and type of eyes-on-road and hands-on-wheel alerts, including the glance behavior and steering control level adjacent to alert onsets (RQ 1-2), and (b) L2 hands-free driving segments specifically to allow for estimation of hands-free driving proportion, alert presence and visual attention specifically occuring when L2 hands-free driving was activated (RQ 3). This data segmentation is illustrated in Fig. 2.

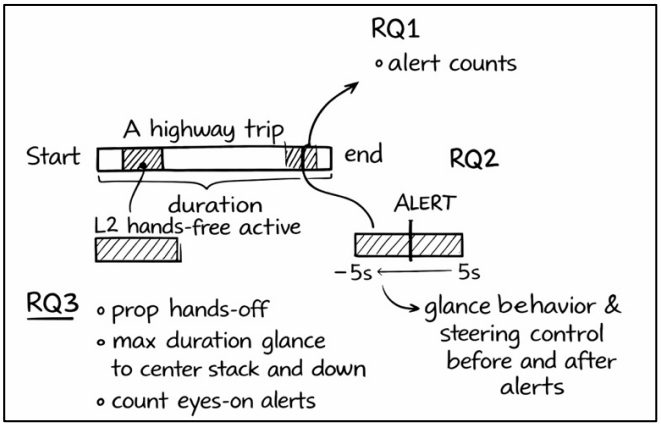


Fig. 2 *Process of extracting sub-segments and relevant variables from highway data*

1.2.1 Data Visualisation and Statistical Analyses

Steering control and glance location proportions were visualised 5s before and after alert onset to assess behavioral shifts associated with alerts. A set of Bayesian general linear models were fit to the dataset to estimate: (a) Alert rates (RQ1) using Negative Binomial models with an offset of segment duration (mins), (b) Response times to alerts (RQ2) using lognormal models, and (c) The correlation between hands-off behaviors and visual attention (RQ3) combining a binomial model of alert presence and a lognormal model of maximum glance durations to down and center stack regions. All models included random intercepts for drivers. Results are reported as posterior summaries (95% credible intervals), transformed to the response scale where appropriate.

2. Results

Across 167 hours of highway driving, we observed 340 first-stage (visual-only) hands-on-wheel alerts, 85 that escalated to visual + audio alerts, and 2 that escalated to resume-control events. For eyes-on-road alerts, 454 first-stage alerts were observed—often initiated directly with visual + audio—of which only 1 escalated to resume control.

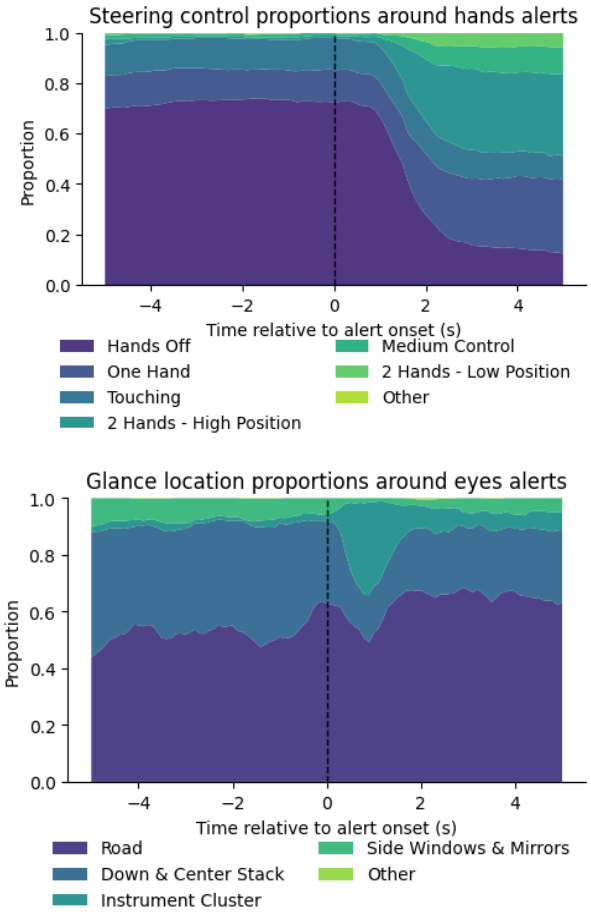


Fig 3. *Proportion steering wheel control level (top) and glance location (bottom) surrounding alerts. These figures (and corresponding models) exclude 0.4% events with missing data for hands and 19.8% events for glance.*

3.1 Rate of alerts per minute of driving

First-stage hands-on alerts occurred at a rate of 1.7 per hour of highway driving (95% CI: 1.0–2.8), with driver variability of 0.74 on the log-rate scale. Eye alerts occurred less frequently, in general, with an estimated rate of 0.9 per hour of highway driving (95% CI: 0.4–2.0), with a driver variability of 1.34 on the log-rate scale.

2.1 *Hands-on and eyes-on response times to alerts*

Alerts were not always triggered when hands were off the wheel or eyes were off the forward roadway (Fig. 3). Hands-on-wheel alerts coincided with hands off the wheel in approximately 78% of cases (95% CI: 51–92%). When a manual response was required, drivers returned their hands to the wheel within a median of 2.1s (95% CI: 1.8–2.5s).

For eyes-on-road alerts, drivers were not looking ahead at alert onset in approximately 38% of cases (95% CI: 24–57%). When a visual response was required, drivers redirected their gaze to the roadway within a median of 0.5s (95% CI: 0.4–0.7s).

2.2 *Correlations between hands-off-wheel and eyes-off-road behaviors*

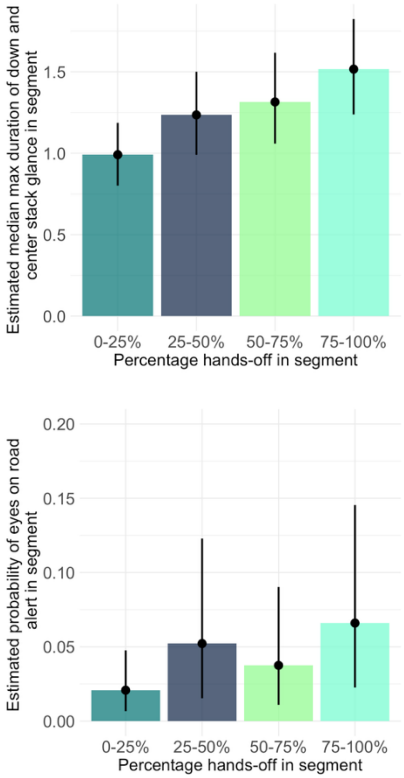


Fig 4*. Estimated median and 95% credible interval limits of maximum durations of glances towards down and center stack regions within L2 hands-free segments (top) and estimated median and 95% credible probability of observing an eyes-on-road alert within a L2 hands-free segment (bottom).*

When down/center-stack glances were present, greater hands-off time within a segment was associated with longer maximum glance durations (Fig. 4, top; Table A3). Compared with the baseline (0–25% hands-off), all higher hands-off intervals (25–50%, 50–75%, and 75–100%) showed 95% CrIs entirely above zero, indicating increased maximum glance duration. Hands-off time was also associated with the likelihood of an eyes-on-road alert (Fig. 4, bottom; Table A4), but only when contrasting the lowest, second lowest, and highest hands-off categories: segments with 25-50% and 75–100% hands-off time were more likely to trigger an eyes-on-road alert than those with 0–25%.

3. Discussion

Novice users of hands-free L2 automation experienced about one to two alerts per hour of highway driving. Between-driver variability was greater for eyes-on-road alerts, indicating that gaze-based monitoring is more sensitive to individual differences, whereas hands-on-wheel alerts reflect both driver behaviour and driving context.

Alerts also occurred when drivers were compliant, suggesting conservative timing, sensing uncertainty, or context-dependent alert logic. When hands were off at alert onset, drivers restored contact within a median of 2.1 s, whereas gaze returned to the roadway in about 0.5 s, consistent with reported differences between visual and manual components of driver re-engagement (Gershon et al., 2023).

Alert onset coincided with increased instrument-cluster glances, similar to responses to take-over requests in real traffic (Pipkorn et al., 2023, 2024). This suggests that alerts may prompt a brief system-state check via the HMI before attention returns to the roadway.

Greater hands-off driving was associated with longer off-road glances and a higher likelihood of eyes-on-road alerts, suggesting that reduced manual control is accompanied by shifts in visual attention allocation. Future work should determine whether this reflects increased engagement in non-driving tasks.

4. Conclusions

Novice users of a commercial hands-free L2 system experienced on average 1–2 alerts per hour, with hands-on alerts occurring about twice as often as eyes-on alerts. Glance alerts showed greater inter-individual variability, suggesting that visual attention management is more sensitive to individual differences than manual control. Drivers responded relatively quickly to alerts, indicating that the system generally re-engaged them effectively. However, the frequent occurrence of alerts when drivers were already compliant suggests limitations in sensing, alert timing, or algorithmic sensitivity that may reduce system credibility over time. Finally, the association between increased hands-off driving and longer off-road glances suggests that drivers who used the hands-free system more extensively also spent more time looking away from the forward roadway; future research should examine the benefits and risks of this pattern.

Acknowledgments

Data and analysis utilized in this study were supported by the Advanced Vehicle Technologies (AVT) Consortium at MIT (http:// agelab.mit.edu/avt). The first and second authors' efforts in the development of this publication were supported by the New England University Transportation Center (NEUTC; US DOT UTC Grant number 69A3552348301). The views and conclusions expressed are those of the authors and have not been sponsored, approved, or endorsed by sponsoring parties.

References

Cadillac. (2025). *Super Cruise*. https://www.cadillac.com/technology/super-cruise

Ford. (2025). *BlueCruise is Hands-free Highway Driving*. https://www.ford.com/technology/bluecruise/

Gershon, P., Mehler, B., & Reimer, B. (2023). Driver response and recovery following automation initiated disengagement in real-world hands-free driving. *Traffic Injury Prevention*, *24*(4), 356–361. https://doi.org/10.1080/15389588.2023.2189990

Gershon, P., Seaman, S., Mehler, B., Reimer, B., & Coughlin, J. (2021). Driver behavior and the use of automation in real-world driving. *Accident; Analysis and Prevention*, *158*, 106217. https://doi.org/10.1016/j.aap.2021.106217

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2019). The "Out-of-the-Loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work*, *21*(1), 87–98. https://doi.org/10.1007/s10111-018-0525-8

Morando, A., Gershon, P., Mehler, B., & Reimer, B. (2021). Visual attention and steering wheel control: From engagement to disengagement of Tesla Autopilot. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 1390–1394. https://doi.org/10.1177/1071181321651118

Pipkorn, L., Dozza, M., & Tivesten, E. (2024). Driver Visual Attention Before and After Take-Over Requests During Automated Driving on Public Roads. *Human Factors*, *66*(2), 336–347. https://doi.org/10.1177/00187208221093863

Pipkorn, L., Tivesten, E., Flannagan, C., & Dozza, M. (2023). Driver Response to Take-Over Requests in Real Traffic. *IEEE Transactions on Human-Machine Systems*, *53*(5), 823–833. IEEE Transactions on Human-Machine Systems. https://doi.org/10.1109/THMS.2023.3304003

Pipkorn, L., Zach Noonan, T., Mehler, B., Reimer, B., & Gershon, P. (2025). Changes in glance and hands-on-wheel behaviors when engaging with non-driving-related tasks while using Tesla Autopilot. *Transportation Research Part F: Traffic Psychology and Behaviour*, *109*, 232–241. https://doi.org/10.1016/j.trf.2024.11.004

SAE International. (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. https://www.sae.org/standards/content/j3016_201806/

Victor, T. W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., & Ljung Aust, M. (2018). Automation Expectation Mismatch: Incorrect Prediction Despite Eyes on Threat and Hands on Wheel. *Human Factors*, *60*(8), 1095–1116.

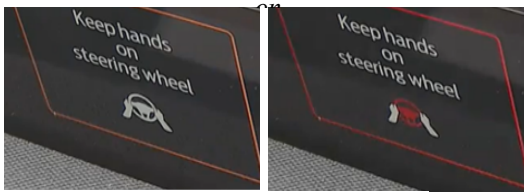
Appendix A

A.1 Images of different alert types

First stage watch the road alerts



First stage (left) and Escalated (right) hands-on-wheel alerts triggered during BlueCruise hands-off



Resume control



A.2 Glance and steering wheel control annotations

Tab. A1 Glance-annotations

Road	any glance directed outside the windscreen (e.g., toward the forward roadway)
Instrument cluster	any glance to the instrument cluster or the steering wheel region
Down and center stack	any glance to the center stack, the in-car touchscreen, or down when looking at a smartphone or other object in the lap.
Side Windows and Mirrors	any glance to the side windows or mirrors (incl. rearview mirror)
Other	any glance not belonging to any other category
Not available	if glances were temporarily not visible due to lighting conditions.

Tab. A2 Hands-annotations

High – two hands	if both hands were on the steering wheel at the optimal 3–9o'clock position
Medium	if the driver had only one hand at the 3–9o'clock position and the other hand elsewhere on the steering wheel *Low – two hands:* if both of the driver's hands were not at the 3–9o'clock position
Low – two hands	if both of the driver's hands were not at the 3–9o'clock position
One hand	if only one hand was placed on the steering wheel regardless of its position
Hands off	if no hand held the steering wheel
Touching	when a finger touched the wheel
Other	when the hands placement did not fit into any of the other categories (e.g., hands hoovering)
Not available	if hands were temporarily not visible due to lightning conditions.

A.3 Model tables

Tab. A3 Maximum duration down-and-center stack glances model

Parameter	Estimate	SE	95% CrI
$\beta_{0-25\%}$	−0.01	0.10	[-0.22, 0.17]
$\beta_{(25-25\%)-(0-25\%)}$	0.22	0.05	[0.11, 0.33]
$\beta_{(50-75\%)-(0-25\%)}$	0.29	0.05	[0.18, 0.39]
$\beta_{(75-100\%)-(0-25\%)}$	0.43	0.05	[0.34, 0.52]

Tab. A4 Glance alert model

Parameter	Estimate	SE	95% CrI
$\beta_{0-25\%}$	−3.98	0.52	[−5.01, −3.00]
$\beta_{(25-25\%)-(0-25\%)}$	0.95	0.28	[0.38, 1.47]
$\beta_{(50-75\%)-(0-25\%)}$	0.59	0.35	[−0.10, 1.24]
$\beta_{(75-100\%)-(0-25\%)}$	1.22	0.29	[0.65, 1.78]
$\beta_{segment_duration}$	0.67	0.06	[0.55, 0.79]

Reducing Ambiguity in Emergency Vehicle Interaction: Strengthening Communication with LED Displays

Kajsa Weibull [1*], Elinor Lundgren [2], Björn Lidestam[1], Magnus Nylin [1]

[1] Swedish National Road and Transport Research Institute, Linköping, SE-581 95 Linköping, Sweden
[2] Department of Computer and Information Science, Linköping University, SE-581 83 Linköping, Sweden
[*]kajsa.weibull@vti.se

Abstract:
During emergency driving and at accident scenes, communication to surrounding road users must be clear and unambiguous to promote a safe traffic environment. In Sweden this is done using emergency vehicle lighting (primarily flashing blue or red/blue) and sirens. However, as shown in earlier studies, it is not always clear what behavior is expected, and the effect of auditory communication may be reduced in modern cars by good insulation and the use of different kinds of auditory media. A possible complement, though currently not allowed to be used operationally in Sweden, is the addition of a LED-display with supplementary text messages. In this study, it was explored how personal vehicle drivers would perceive such a solution. Interviews with emergency vehicle drivers were conducted to identify relevant scenarios, the readability of the LED-display was investigated, and an online survey (*N*=2265) was conducted. The participants were presented with video recordings showing four different scenarios: pull over – stop signal, roadside accident, pull over – take the next exit, and emergency driving – calling for clear way, and were randomly assigned to see emergency vehicle lights only, LED-display only, or a combination. For each situation, questions about anticipated driver actions, perceived certainty, perceived visibility, perceived stress, perceived danger, and perceived urgency were answered. The results showed that the use of the LED-display, especially together with emergency vehicle-lights, provides a more effective way of promoting desired behavior and probably reducing stress. Future studies should complement these findings by driving experiments.

1. Introduction

Communication with other road users during emergency driving is crucial for safe and efficient operations. However, sirens may not be heard in modern, well-insulated cars and in competition with other sound sources, and previous studies have shown that light signals currently used are not always easy to interpret for other road users (Weibull et al., 2024a). One solution is to use complementary LED text signs with text adapted to different situations, but current Swedish legislation, contrary to some other countries, does not allow that. This study was performed by VTI with support from Standby AB to explore how this may work in a Swedish context. The main part of the study was an online survey to find out how drivers thought they would react in different traffic situations when encountered with only light signals, only the LED text sign, or a combination of both.

2. Method

To decide on relevant scenarios for the online survey, two pre-studies were performed. The first was an interview study with eight emergency operation professionals. The aim of the interview study was to identify relevant scenarios where they experienced difficulties in communication with surrounding drivers. In the second pre-study, the readability of the LED-displays was assessed. The results indicated that the LED-display was readable from about 100 meters. The participants in the pre-studies were all participating voluntarily, informed about the study, and gave informed consent to participate. Research IDs were used, and identifying information was removed from the material. For the online survey, participants were recruited via Facebook advertising. The participants were anonymous. The respondents had the opportunity to email questions to the responsible researcher.

The participant was randomly chosen to assess videos with emergency vehicle lights only, LED-display only, or a combination of both. Regardless of communication variant, all participants watched the same scenarios: pull over – stop signal, roadside accident, pull over – take the next exit, and emergency driving – calling for clear way, identified in the first part of the study. Hence, each participant watched four videos. After each video, they answered questions about how the demonstrated communication would affect:

- Traffic behaviors
- Perceived urgency
- Perceived visibility
- Perceived certainty
- Perceived danger
- Perceived stress

Traffic behavior was assessed by choosing from a list of possible actions, inspired by Ledin (2022). The possible actions were:

- Step on the brake
- Pull over to the side of the road
- Stop

- Accelerate
- Ease off the accelerator
- Use the turn signal
- Use the hazard lights
- Take the next exit
- Continue at the same speed and lateral position

3. Results

A total of 2,265 drivers (74% men, 25% women, 1% preferred not to say) responded to the survey. They were on average 60 years old (*SD* = 13.4). Odds ratio (OR) analyses were performed to determine how a LED-sign would affect reported driving behavior. An odds ratio of 1.0 would indicate that the alternative communication variant did not change the reported behavior, compared to continuing driving without changing the behavior (Baseline). An odds ratio below 1.0 would indicate a decreased likelihood of an action, while an odds ratio above 1.0 would indicate an increased likelihood of an action. Confidence intervals of 95% were used to determine statistical significance.

3.1 Pull over: Stop signal

The LED-sign displayed the message "*STOPP*" (Eng: *STOP*) (Figure 1). A combination of LED and current lights nearly doubled the likelihood of stopping (OR = 1.94) compared to baseline. Only blue lights were associated with pull over to the side of the road to a higher extent (OR = 4.21) compared to baseline.



Fig. 1 *– Displaying the Stop signal scenario*

3.2 Roadside accident

The LED-display stated, "*Olycka: Sänk farten*" (Eng: *Accident: Reduce speed*) (Figure 2). Blue lights decreased the likelihood of stopping (OR = 0.08) and increased the likelihood of pulling over to the side of the road (OR = 1.98). The LED-display with the current lighting increased the likelihood of braking (OR = 1.26) and decreased the odds of pulling over (0.65). A similar effect was observed with the LED display without lights (Braking OR = 1.34, Pulling over OR = 0.59).



Fig. 2 *– Displaying the Roadside accident scenario*

3.3 Pull over: Take next exit

The LED-sign displayed the message "*TA NÄSTA AVFART*" (Eng: *TAKE the NEXT EXIT*) (Figure 3). The LED-display and combination of LED and lights increased the probability of the respondent taking the next exit (LED OR = 14.17; *Combination* OR = 7.01), while the blue lights decreased the likelihood (OR = 0.39). Instead, the blue lights increased the likelihood of pulling over to the side of the road (OR = 1.85) and stopping (OR = 2.54), which was not the desired outcome.



Fig. 3 *– Displaying the Stop: Take next exit*

3.4 Emergency driving

The LED-sign displayed the message "*HÅLL FARTEN*" (Eng: *MAINTAIN SPEED*) (Figure 4). The desired behavior was that the driver would continue driving with the same speed and lateral position. The LED-display and combination increased the likelihood of maintaining speed and lateral position (LED OR = 24.18; *Combination* OR = 10.06). The blue lights increased the probability of pulling over (OR = 2.39), while the LED-display and combination decreased the likelihood (LED OR = 0.03; *Combination* OR = 0.12).



Fig. 4 *– Displaying the Emergency driving scenario*

3.5 Perceived Urgency, Visibility, Certainty, Danger and Stress

The respondents indicated their experienced level of urgency, visibility, certainty, danger, and stress on a level of 0-100 (0 = Not at all, 100 = extremely, e.g., urgent). To examine the differences pairwise comparisons were conducted. The blue-red lights generated the highest sense of urgency in all scenarios where that light combination was included (scenarios 1, 3, and 4). In all scenarios, the blue lights or the combination of LED and lights generated the highest scores for visibility. In the emergency driving scenario, the LED and combination variant received significantly higher scores of certainty compared to blue lights. The level of danger was ranked higher for blue lights in scenarios 3 and 4, compared to LED and combination. The perceived level of stress was, in general, higher in scenarios 3 and 4, compared to scenarios 1 and 2. More specifically, the highest level of stress was reported when the respondent was presented with blue lights, followed by a combination of lights and the LED display. The LED display without lights was reported as the least stressful alternative in scenarios 3 and 4.

4. Discussion

This study shows how LED-displays can be used as a complement to today's emergency vehicle lighting. The use of an LED-display, both with and without the combination of the current emergency vehicle lights, increased drivers' likelihood of performing as desired, such as pulling over in Scenario 1, braking in Scenario 2, taking the next exit in Scenario 3, and continuing driving in Scenario 4. In general, drivers are aware of what is expected from them in a standard "move over" scenario (Weibull, 2024a), but when something else is expected from them in an interaction with an emergency vehicle, they can become insecure and stressed (Weibull, 2024b). Therefore, the use of an LED-display has the potential to support drivers when encountering emergency vehicles in traffic

Future studies should consider how drivers' attention is affected when provided with LED-displays in an emergency vehicle interaction.

5. Conclusions

LED-displays have great potential as a complementary communication method for emergency vehicles. However, the current Swedish legislation does not allow emergency vehicles to display such information.

6. Acknowledgments

The study was supported by Standby Sweden, part of Standby Group.

7. References

Ledin, J. (2022). Utformning av blåljus påverkar trafikbeteenden: Alternative Emergency Vehicle Lighting Affects Traffic Behaviors [Master Thesis]. Linköpings University.

Weibull, K., Lidestam, B., Holm, J., & Prytz, E. (2024a). Alternative Emergency Vehicle Lighting Affects Traffic Behaviors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 68*(1), 1-7. doi:10.1177/10711813241260291

Weibull, K., Lidestam, B., & Prytz, E. (2024b). Upplevda svårigheter vid interaktion med utryckningsfordon och hur EVA-varningar kan underlätta. In F, Hellman and M, Haraldsson. (Eds.), *Sammanställning av referat från Transportforum 2024* (p. 299–300). Swedish National Road and Transportation Research Institute.

Baseline gaze behavior of younger and older drivers under undistracted conditions in cyclist intersection conflicts

Yuqing Zhao [1], Koki Hoshino [1], Jordanka Kovaceva [2*], Sarang Jokhio [2], Koji Mizuno [1], Robert Thomson[2]

[1] Nagoya University, Foro-cho, Chikusa-ku, Nagoya 464-0803, Japan
[2] Chalmers University of Technology, SE-412 96 Göteborg, Sweden
[*] jordanka.kovaceva@chalmers.se

Abstract: This study aims to understand drivers' gaze behavior across age groups during normal driving. Thirty-six younger drivers and 24 older drivers participated in a driving simulator experiment involving two roads of different widths. The results showed that both groups focused more on the near side, and older drivers' gaze was closer to the ego vehicle. Drivers scanned more frequently when approaching intersections, and this behavior was more prounced toward the near side on the narrower road. This study may provide insights into the optimization of driver mornitoring systems based on the age to enhance the road safety.

1. Introduction

Understanding baseline driver behavior is essential for evaluating the safety impact of advanced driver assistance systems (ADAS), and automated driving systems (ADS), and for developing and validating driver monitoring systems (DMS). In particular, reliable baseline gaze behavior from attentive, non-distracted drivers is required to assess deviations caused by distraction, fatigue, or automation-induced states such as reduced monitoring or out-of-the-loop behavior. Such baseline data are also necessary for parameterizing driver gaze behavior models used in safety impact simulations. However, limited empirical data exist on how baseline visual behavior varies across driver demographics, especially in safety-critical situations involving vulnerable road users (VRUs).

This study investigates age- and road-width-related differences in driver gaze behavior using a driving simulator experiment.

2. Method

2.1 Driving simulator and scenario setup

A simulator based on the CARLA open-source platform (v0.9.15) was used, with road implemented through the MATLAB RoadRunner. Participants controlled the vehicle using a Logitech G27 steering wheel and pedal set, while the driving scenarios were displayed on a 40-inch Samsung LED screen. Drivers' gaze and head movements were recorded using a Logitech MX Brio webcam and analyzed with the OpenFace 2.0 toolkit (Baltrušaitis et al., 2018) to estimate gaze direction and facial parameters.

To avoid fatigue or distraction caused by long driving, two short-length urban scenarios with cyclist crossing events were developed. In Scenario A, drivers traveled along a 1081 m road with five intersections. The road width was 12.1 m, with two 3.25 m lanes. Scenario B involved a 3136 m roadway with ten intersections. The road width was 15.9 m, with two 4.05 m lanes. A cyclist emerged from left at 15 km/h at the final intersection in Scenario A and at the ninth intersection in Scenario B. Drivers were instructed to maintain a target speed of 50 km/h and drive as they normally would. The entire driving, including practice course, lasted less than 10 minutes.

2.2 Participants

Sixty valid Swedish or EU licensed drivers participated, including 36 younger drivers aged under 65 years (mean age 34.5 years) and 24 older drivers aged 65 years and above (mean age 70.3 years), all without diagnosed eye diseases. Participants completed both scenarios in a randomized order. This study was approved by the Ethical Review Committee of Chalmers University of Technology.

Drivers' gaze distributions were analyzed using Kernel density estimation (KDE) and plotted in heatmaps, and scan behavior was analyzed using Glance Algorithm (Swan et al., 2021) in MATLAB.

2.3 Statistical analysis

Since each participant generated multiple scans, linear mixed models (LMMs) were used to examine the fixed effects of age, distance to the intersections, scan direction (left/right), and lane width across both scenarios on gaze behavior. The significance level was set at 0.05, and all statistical analyses were conducted using R 4.4.2.

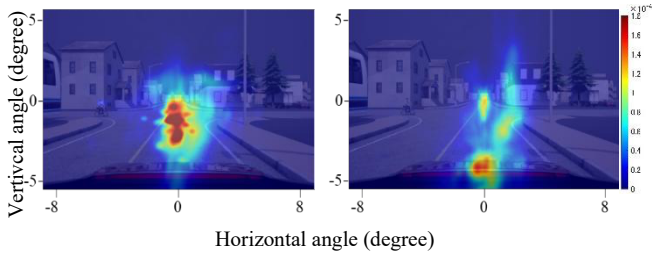
3. Results

3.1 Gaze distribution

Younger and older drivers' gaze distributions in Scenarios A (3.25 m lanes) and Scenario B (4.05 m lanes) were plotted as heatmaps, are shown in Figure 1 and Figure 2, respectively. Older drivers' gaze was distributed closer to the ego vehicle compared to that of younger drivers in both scenarios. Particularly, in the narrower Scenario A, older drivers gazed closer to the speedometer displayed at the bottom middle of the screen.

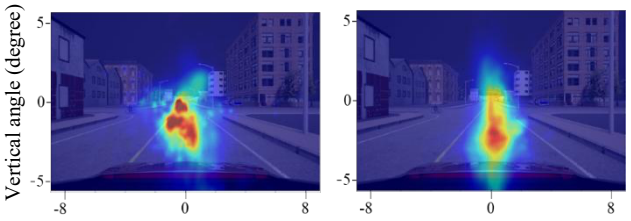
In the horizontal direction, the gaze of younger drivers was more widely distributed than that of older drivers. The average median gaze angles (representing where drivers gazed most) of younger and older drivers in Scenario A were 0.47° and 0.64°, respectively. The average median gaze angle of younger and older drivers in Scenario B was 0.26° and 0.28°, respectively. Both younger and older drivers directed

their gaze to the right side (near side) in both scenarios. However, the LMM results showed that age road width did not significantly affect the median horizontal gaze angle (Age: $F(1, 44.7) = 0.016$, $p = 0.899$; Road width: $F(1, 38.4) = 1.565$, $p = 0.219$).



a) Younger drivers *b) Older drivers*

Figure 1: Gaze distributions of younger and older drivers in Scenario A with lane width 3.25 m



a) Younger drivers *b) Older drivers*

Figure 2: Gaze distributions of younger and older drivers in Scenario B with lane width 4.05 m

3.2 Scan behaviour

Age did not significantly affect drivers' scan frequency when approaching intersections (F (1, 46) = 2.953, p = 0.092). Other significant parameters are shown in Figure 3. Drivers scanned more frequent when approaching intersections, but their scan frequency decreased scans within 20 m of intersections. However, in the narrower scenario, drivers' scanned towards right side more often, especially within 20 m of the intersections (Lane width: $F(1, 906) = 6.734$, $p = 0.010$; Scan direction: $F(1, 906) = 4.935$, $p = 0.027$).

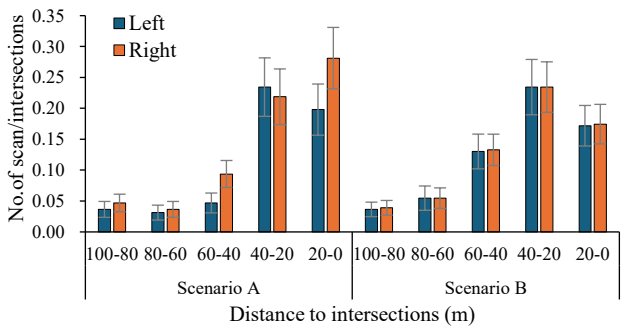


Figure 3: Relationship between scan frequency, distance to intersections, and scan direction in Scenario A and B.

4. Discussion

Drivers showed different gaze behaviors across age groups during normal driving. Older drivers checked the speedometer more frequently than younger drivers, which was more pronounced on the narrower road. This may be explained by the fact that controlling vehicles was more difficult for older drivers under tasks, or that they adopted different task-prioritization strategy.

Both age-group drivers tended to scan the near side more often, which was more obvious on the narrower road. This may suggest that drivers regard the near side as more dangerous, expecting hazards from it,or this may be due to the right-of-way rule when approaching intersections.

Since the scenarios were shown on a single display, which may have limited drivers' head movements, the results differed from previous studies showing that older drivers scan less frequently (Savage et al., 2020). Under the heavy cognitive load, drivers' gaze focuses more on the center of the road (Engström et al., 2005). Whether the narrower gaze distribution and lower scan frequency observed in older drivers are due to cognitive load or physical decline should be investigated in the future.

This study provides baseline gaze behavior of attentive drivers across age groups in different road width scenarios. Such baseline data are essential for developing and validating DMS-based ADAS and ADS, and calibrating driver models used in virtual safety impact assessments (EuroNCAP, 2023; 2025). Future research should extend this work to additional traffic environments and examine how gaze behavior changes under distraction, cognitive load, and interactions with automated driving systems.

5. Conclusions

The results contribute to understanding individual differences in driver behavior and provide baseline gaze patterns that can support the evaluation of inatention information systems based on DMS and driver behavior models used in safety impact assessments of ADAS and ADS.

6. Acknowledgments

This study was supported by GENIE program and Transport Area of Advance project TREND.

References

Baltrušaitis, T., Zadeh, A., Lim, Y., Morency, L. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit, IEEE Int. Conf. Autom. Face Gesture Recognit.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. Transp. Res. Part F: Traffic Psychol. Behav. 8(2), 97-120.

Savage, S. W., Zhang, L., Swan, G., Bowers, A. R. (2020). The effects of age on the contributions of head and eye movements to scanning behavior at intersections. Transp. Res. Part F: Traffic Psychol. Behav., 73, 128-142.

Swan, G., Goldstein, R. B., Savage, S. W., Zhang, L., Ahmadi, A., Bowers, A. R. (2021). Automatic processing of gaze movements to quantify gaze scanning behaviors in a driving simulator. Behav. Res. Methods, 53(2), 487-506.

Euro NCAP. (2023). *Euro NCAP rating review 2023*. https://www.euroncap.com/media/79890/euro-ncap-rating-review-2023.pdf

Euro NCAP. (2025). *Safe driving: Driver engagement* (Version 1.1).https://www.euroncap.com/media/91704/euro-ncap-protocol-safe-driving-driverengagement-v11.pdf

Characterizing and Detecting the Driver Intended Head Movements During attentional and inattentional behaviours: An Open Road and Test Track Study.

Aimé Tiadi[1], Mohamed-Zaki Chellali[1], Pascal Quentin[1], Jean-Marc Tissot[1]
[1] Valeo, Brain Division, 6 Rue Daniel Costantini, Créteil, France
*aime.tiadi@valeo.com

Abstract: Head movements are well known to be strong cues to characterize driving task engagement. When added to eye gazes, both sensory (visual) and motor (head) involvements lead to accurate driver attention assessment. However, distinguishing between intended head movements and unintended ones remains a challenge. This study presents a technical framework for characterizing and detecting head-movement intent. 34 subjects participated in the study. More than 51 hours of driving data were collected on open road and on test track using head/eye tracking camera. Additional head and eye movement scenarios were acquired in lab car. The analysis of movement dynamics, head and gaze alignment, lead to achieve the classification that distinguishes between attentional, inattentional behaviours. We suggest that, these findings can significantly enhance next generation driver monitoring system robustness under real-world conditions.

1. Introduction

When the driver head orientations change deliberately and are associated to the visual shifts towards a given location, both motor and sensory inputs are aligned. This can indicate the driver intent to process relevant information for the driving activity (Zhao et al., 2017); Althoug several studies examined the driver head movements (Lee et al., 2018; Zhang et al., 2019; Khattak et al., 2024; Schakmann et al., 2024), those that dealt with the intended head movements are scarce. Zhang and collaborators (2019) analyzed head and gaze movements before driving maneouvers (e.g lane changes) during naturalistic driving and found that head and gaze yaw are more correlated to driver intention to perform maneuvers. These authors did not applied the study to non driving related tasks and to non intended movements like head drops. Khattak et al. (2024) analysis on the role of driver head pose was related to safety critical events before crashes and near crashes. They found that the angular deviations of the head increased the likelihood of critical events from 4 to 8%.

Our study aims to investigate the driver head and gaze movements to develop and test an algorithms capable to classify between intended and unintended head movements in order to robustify the recognition of driver attention, fatigue and situation awarness and spoofing for next generation of driver monitoring system. Both open road and test track driving sessions data with driving and non driving related tasks were analyzed. This was followed by in vehicle tests to evaluate the detection of the two kinds of movements by the algorithms. Our hypothesis was a weaker detection of inattention head behaviours like head drops (unintended head movements).

2. Method

2.1 Participants

The study included regular 34 subjects having valid driving license with 50% males and 50% females. Their age ranged between 20–65 years old (average: 41.24 years; SD: 16.59),.

The participants were recruited with the help of medical experts located in the south-west of France; neither sleep disorders nor medical advice preventing participants from driving were mentioned and documented; if so, they were excluded from the study. All participants filled and signed a written informed consent document prior to experiments. The local Ethics Committee approved the study.

2.2 Procedure

Each participant performed open road driving session that lasted 90 min. Additional 90 min test track driving sessions were also performed during which the participants runned driving and non driving related tasks in different gaze locations with owl and lizard movements as specified by ENCAP 2023 Driving engagement protocol. Head/eye tracking camera was used for both open road and test track driving sessions. 513.709 data points were collected and used to analyze and classify head movement types through a dedicated algorithm. Finally, another experiment was runned in lab car when the vehicle was in static position.

3. Results

The movement dynamics were analyzed using head and eye gaze angles in different AOIs (Areas Of Interest). Then, in order to distinguish between intended head movements and unintended ones, spatial and temporal criteria were used: head and gaze movement amplitudes, head and gaze alignment (difference between head and gaze angles and amplitudes), head fixation minimum duration. In addition, the head and gaze tracking quality thresholds were taken into consideration to ensure a good quality of tracking data. This approach leads to develop and validate intended head movements algorithm.

3.1 Head Movement Status

Head and gaze movements were analyzed in different AOIs in order to characterize and define the gaze and head alignement (see Fig. 1 to 4).

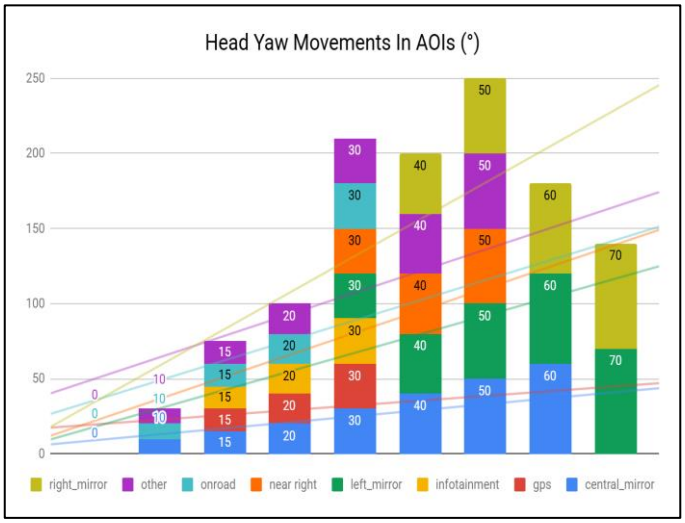


Fig. 1. *Drivers' Head Yaw Movement Distributions in different AOIs with Angle values in degree.*

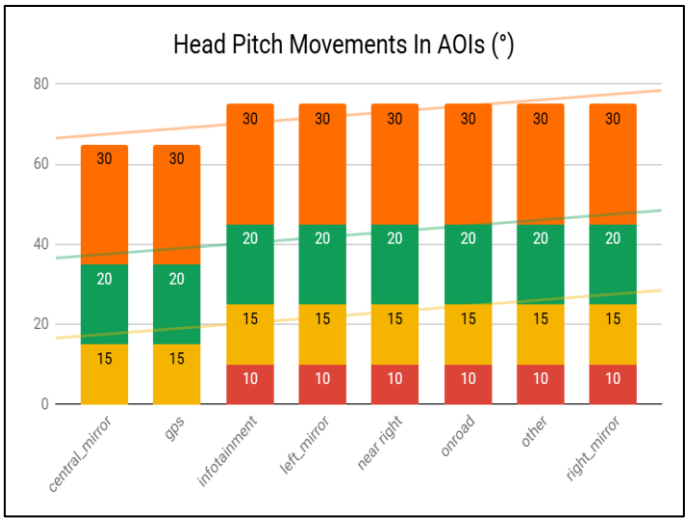


Fig. 2. *Drivers' Head Pitch Movement Distributions in different AOIs with Angle values in degree.*

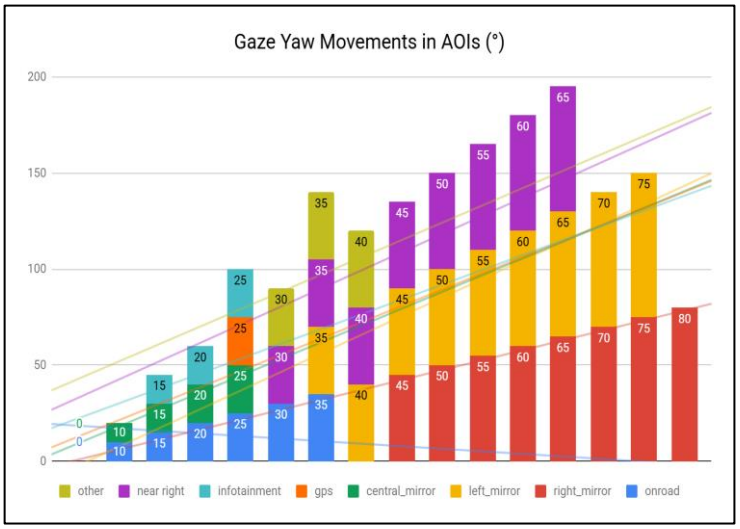


Fig. 3. *Drivers' Gaze Yaw Movements in different AOIs with Angle values in degree.*

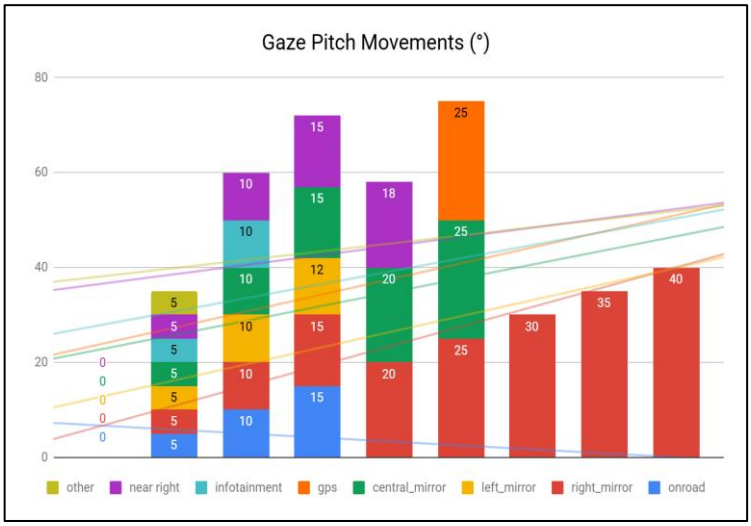


Fig. 4. *Drivers' Gaze Pitch Movements in different AOIs with Angle values in degree.*

3.2 Intended Head Movements Status

3.2.1 On Open Road

By filtering low quality due to steering wheel movements on the driver face that occlude sometimes gaze and head, the detections were achivied on open road driving sessions with a performance that showed an accuracy of 0.97, with 96.55% of sensitivity and 0 false alarm by hour as swhown by the confusion matrix (see table 1).

3.2.2 On Test Track with driving and Non Driving tasks.

In another experiment on test track, the intended head movement approach was applied when the drivers performed driving related tasks and non driving related tasks in different gaze locations with Lizard and owl movements. We expected differences between lizard and owl movements and alignement between gaze and head leading to intend head movements more present in owl movements than lizard ones.

The results from ANOVA showed significant differences between lizard and owl movements when driver performed driving related task by looking on driver side mirror (DSM), F-6.5341310, p= 0.01. Figures 5 and 6 show the intended head movement status for DSM.

The significant differences were also observed between lizard and owl movements for rear view mirror (RVM) F-80.0606342, p= 0.000 and non driving related tasks in Invehicle Infontainment (IVI) F-2541921, p=0.000 as shown on figures 7 to 10.

Table 1. The performance of intended head movement detection for Real driving sessions on open road:

Accuracy	Sensitivity	False Alarm/hour
0.97	96.55	0

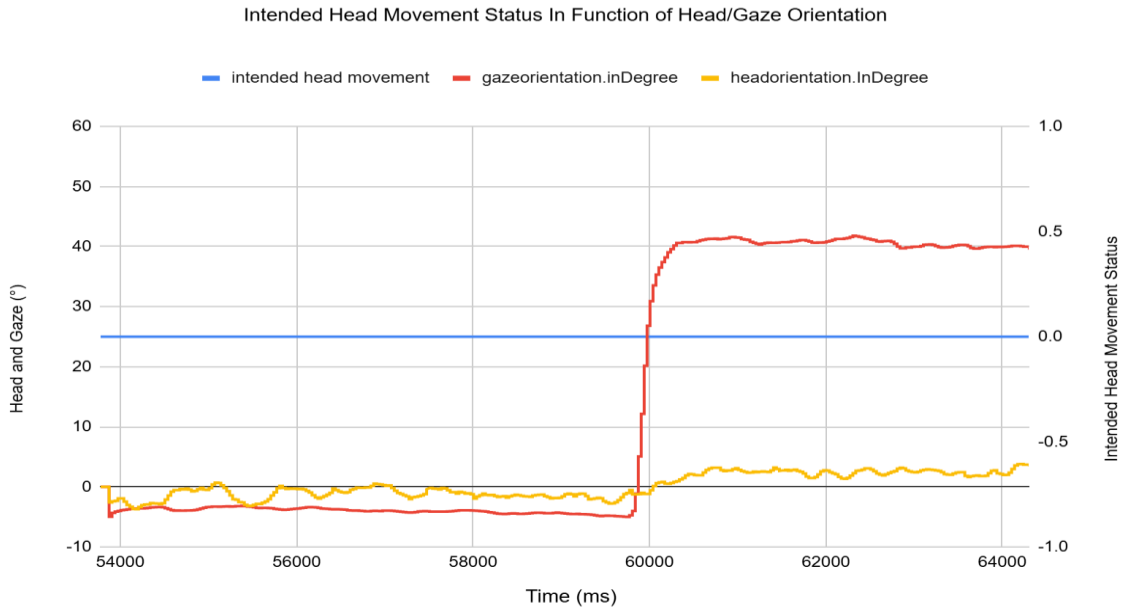


Fig. 5. *Driver intended head movement status for driving related task in Driver side mirror with lizard movements*

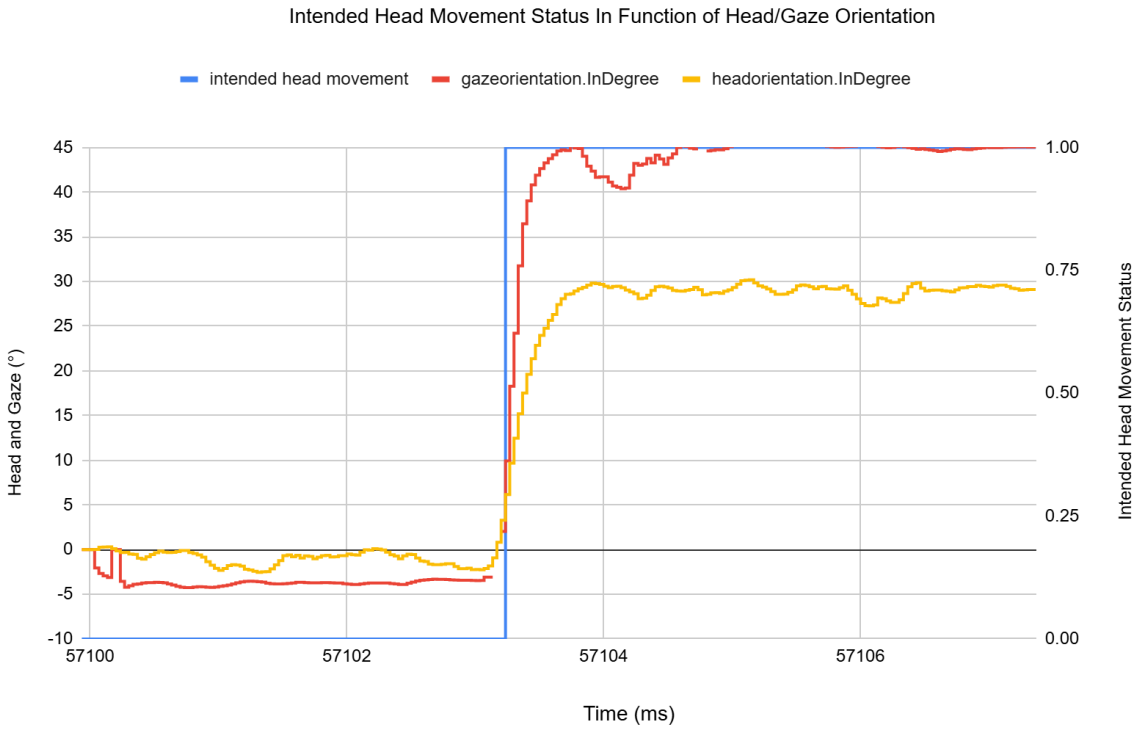


Fig. 6. *Driver intended head movement status for driving related task in Driver Side Mirror with owl movements*

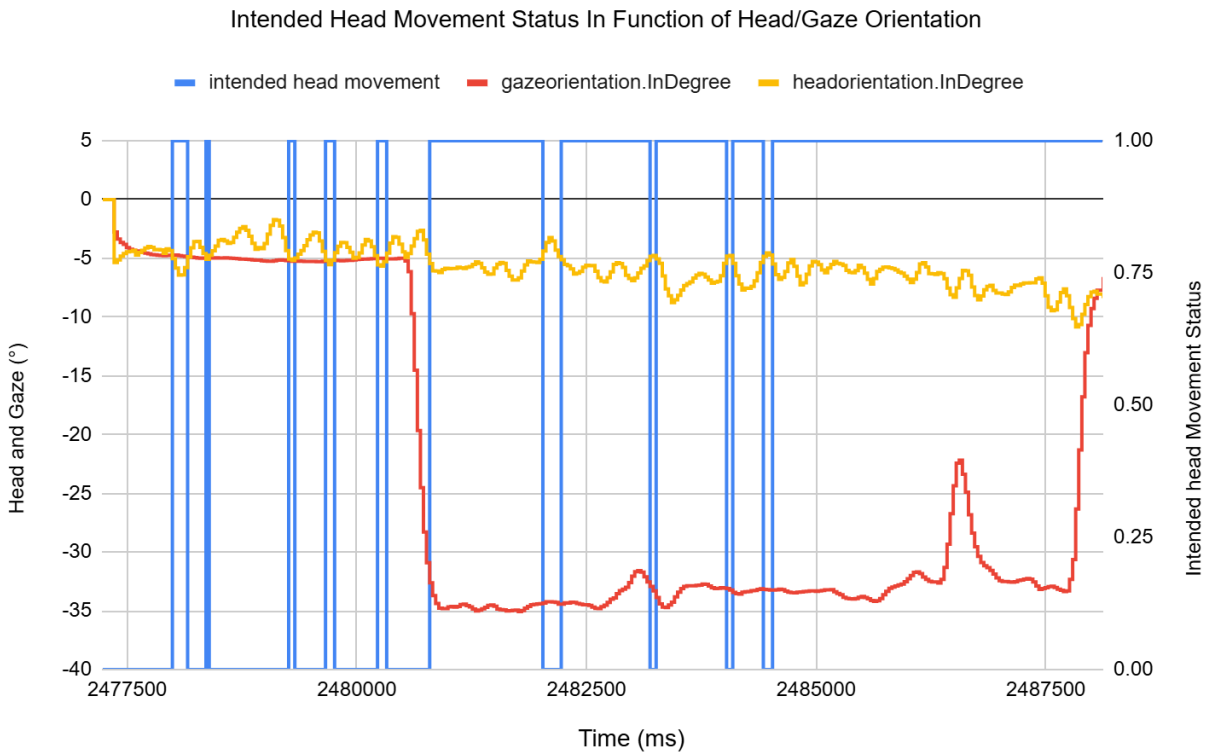


Fig. 7. *Driver intended head movement status for driving related task in Rear View Mirror with lizard movements*

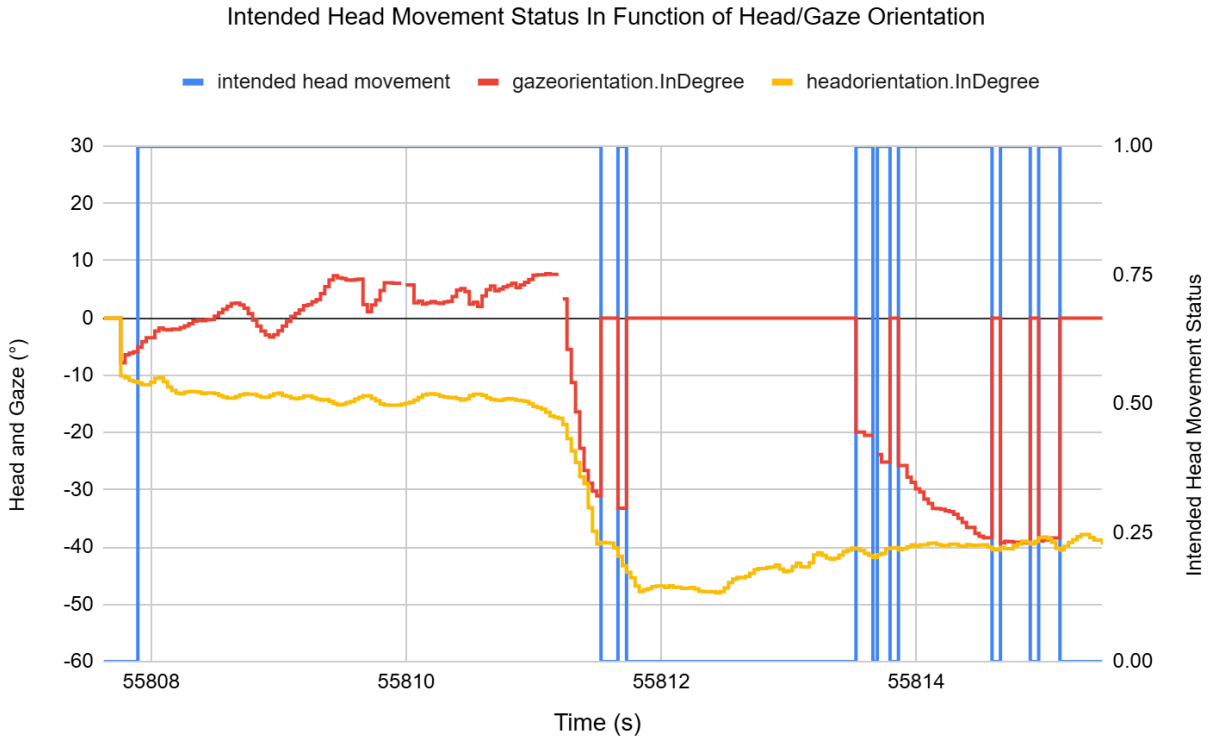


Fig. 8. *Driver intended head movement status for driving related task in Rear view mirror with owl movements.*

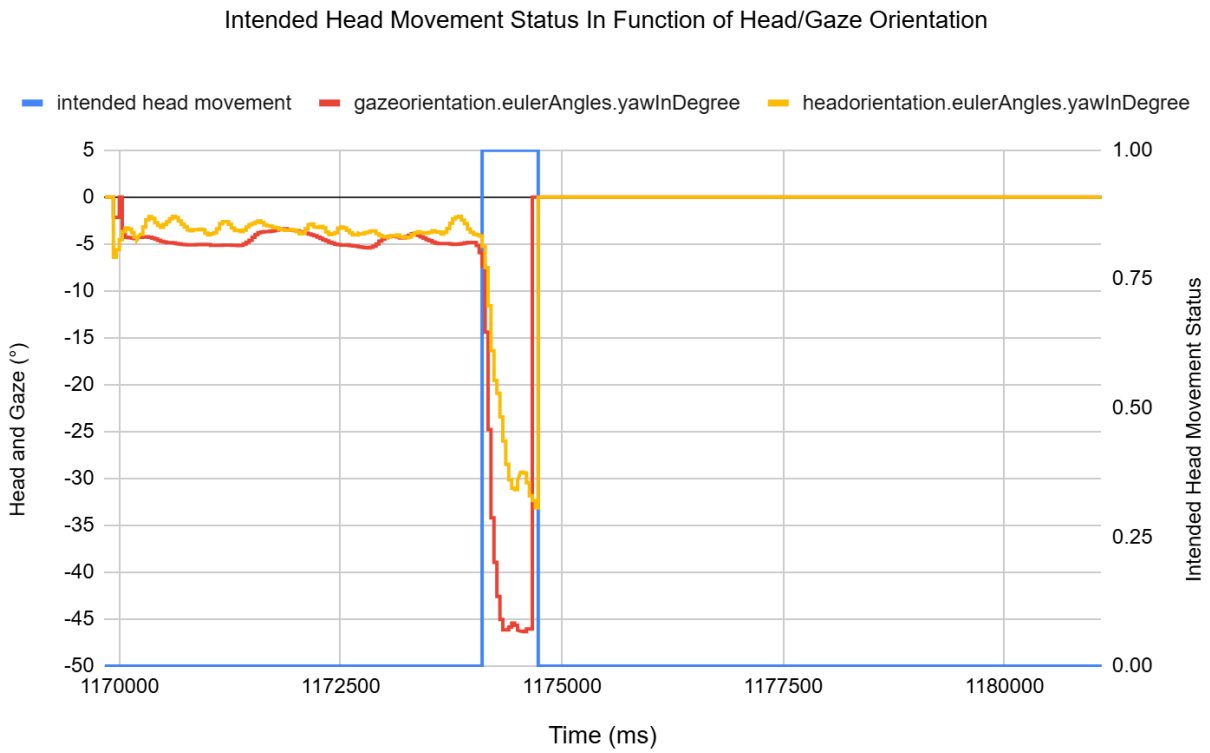


Fig. 9. *Driver intended head movement status for non driving related task in IVI with lizard movements.*

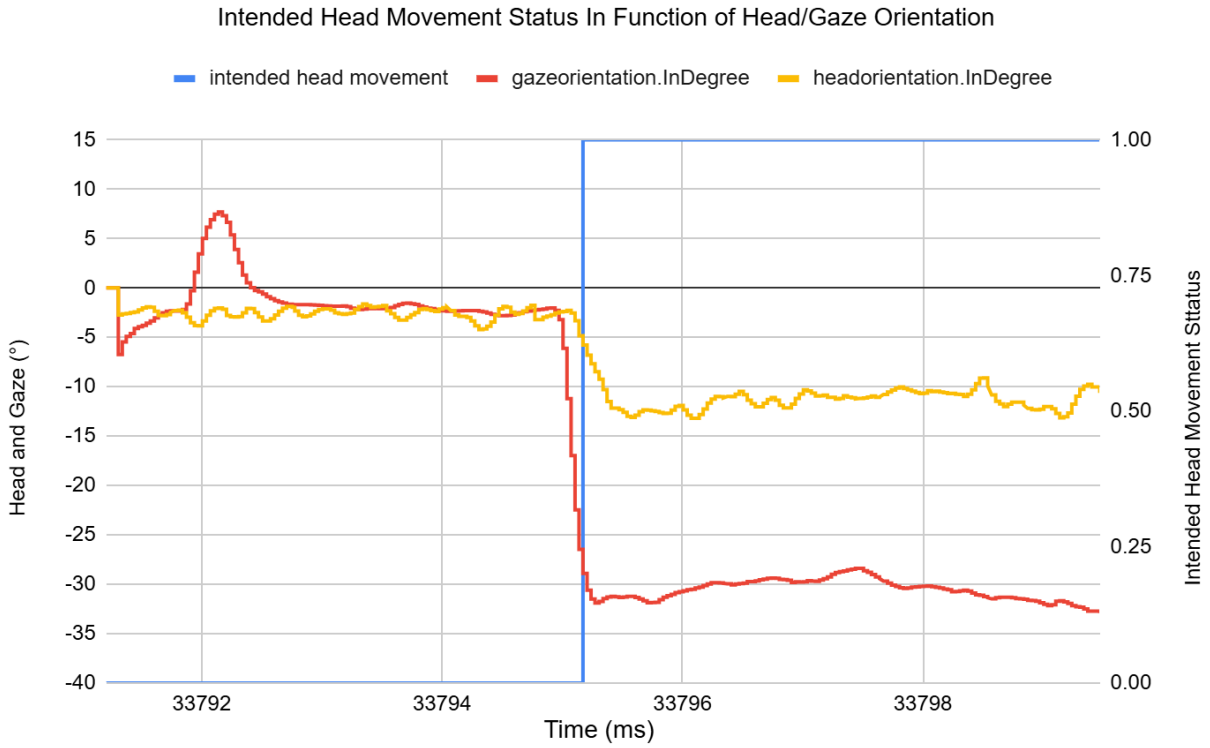


Fig. 10. *Driver intended head movement status for non driving related task in IVI with owl movements.*

3. 3 In Vehicle Tests

The intended head ovement algorithms was also tested, with the embedded version when the vehicle was in static position, while participants performed yaw and pitch movements in order to assess the system's ability to correctly detect and differentiate between attentional vs non attentional behaviour. The yaw movemnts detection rates were 93,75% and 86,67% for large and small amplitude respectively (see Fig. 11). The picth movements detection rates were 97,5% for large amplitude. But we noticed a bad detection rate for small amplitude, 18.42 % (see Fig. 12). As expected also, we observed significant differences between all head movements taken together that are aligned with gaze (Intended behaviours) and head drops related to inattention behaviour: $F(1.8) = 17.88$, $p=0.003$ (See Fig. 13). The detection of intended head movements was significantly greater than inattention head behaviour meaning that algorithms is capable of distinguishing between intended vs unintended head movements.

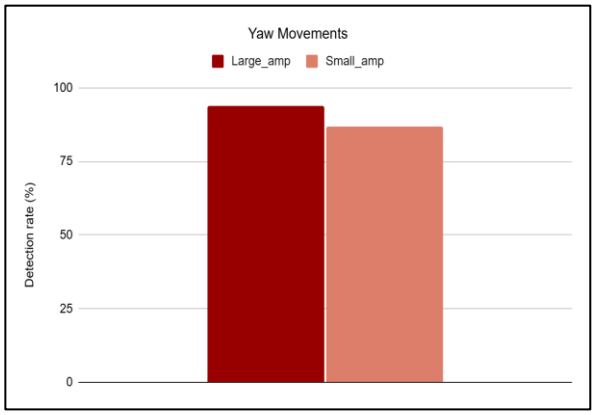


Fig.11 *Detection rate of driver Head/gaze yaw movements when vehicle was in static position.*

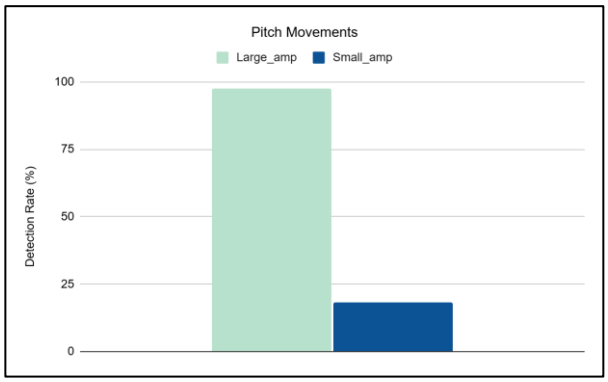


Fig. 12. *Detection rate of driver Head/gaze pitch movements when vehicle was in static position.*

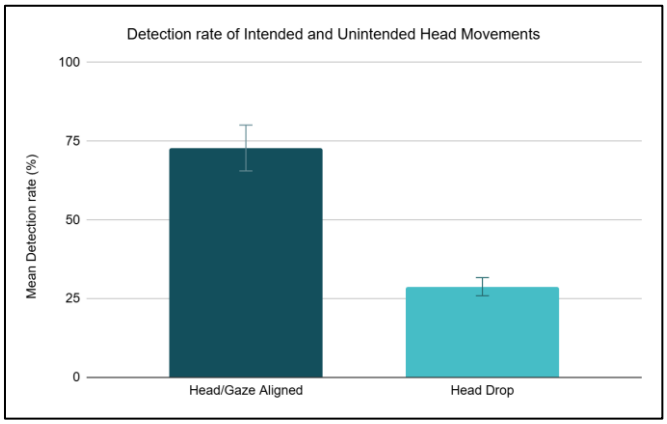


Fig. 13. *Detection rate of driver Intended head movements and inattention behaviour (Head drop, unintended head movements).*

4. Discussion

The results obtained in this study demonstrate a high level of performance of the proposed detection framework under naturalistic open-road driving conditions. After filtering low-quality samples caused by steering wheel movements partially occluding the driver's face, the system achieved an overall accuracy of 0.97, with a sensitivity of 96.55% and zero false alarms per hour. These results indicate that the proposed approach is capable of reliably detecting the targeted intended head movement behaviours in real-world driving environments while maintaining a very low rate of erroneous detections. Lee et al. (2018) previous study has shown that head and gaze correspondence can serve as a reliable proxy for estimating the driver's attentional focus, particularly when eye tracking information is unavailable or unreliable. Moreover, the absence of false alarms per hour represents a particularly important outcome for the proposed intended head movement detection approach.

In addition, as expected our results demonstrate clear differences between the occurences of intended head movements for owl movements compared to lizard ones when drivers performed both driving and non driving related tasks (in DSM, RVM, IVI locations) . This highlights that the proposed detection system supports the fact that intended head movements are more frequently associated with owl-type gaze behavior, where head rotations contribute significantly to visual exploration and show their alignment (Friedman et al., 2016).

Furthermore, the proposed system showed the ability to differentiate between attentional head movements and non attentional ones (head drops). Indeed, despite the limitation for small amplitude pitch movements, for which the detection rate was very low, the statistical analysis demonstrated a clear distinction between intended head movements aligned with gaze behavior and simulated head drops representing inattentive behavior.

This distinction is particularly important in the context of driver monitoring systems, where the ability to differentiate between intentional head movements and signs of driver inattention or fatigue is essential for reliable safety interventions (Fridman et al., 2016; Dong et al., 2011). In real driving scenarios, head movements may reflect normal visual

scanning behaviors required for safe driving, such as checking mirrors or observing the road environment. Conversely, unintended head movements, including nodding or head drops, may signal driver fatigue or reduced alertness. The significant difference observed in this study suggests that the algorithm captures meaningful behavioral cues that allow it to distinguish between these two types of head movements on open road, test track and static position of vehicle.

The relatively poor detection rate observed for small pitch movements indicates that further improvements may be necessary to enhance sensitivity to subtle head orientation changes. Also, further experiments should be applied on a larger dataset.

5. Conclusions

This study shows that the embedded intended head movement detection algorithm performs well for most head movements and is capable of distinguishing between intended head–gaze aligned behaviors and inattentive head drops. The findings highlight the potential of the proposed approach for supporting driver monitoring systems designed to assess driver attention and detect early signs of inattention or fatigue as well as driver situational awarness and non human presence (spoofing).

6. Acknowledgments

We would like to thank our university and medical partners. The authors are very thankful to University and University Hospital Center.

References

Chen, Z., Wu, C., Zhong, M., Lyu, N., & Huang, Z. (2015). Identification of common features of vehicle motion under drowsy/distracted driving: A case study in Wuhan, China. *Accident Analysis and Prevention*, 81, (2015): 251–259.

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, *12 (2)*: 596-614.

European Union Council. (2019). Regulation (EU) 2019/2144 of the European Parliament and of the Council. *Official Journal of the European Union*, L 325: 1- 39.

Fernández, A., Usamentiaga, R., Carús, J., L., & Casado, R. (2016). DriverDistraction Using Visual-Based Sensors and Algorithms. *Sensors,* 16 (1805): 1-44.

Fridman, L., Lee, J., Reimer, B., & Victor, T. (2016). Owl and Lizard: Patterns of Head Pose and Eye Pose in Driver Gaze Classification. *IET Computer Vision*, *10 (4)*: 1-9.

Han, J-H. & Ju, D-Y. (2021). Advanced Alarm Method Based on Driver's State in Autonomous Vehicles. *Ergonomics,* 10 (2796): 1-15.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting Santa Monica, CA: *Human Factors & Ergonomics Society.* (pp. 904- 908).

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati(Eds.), Human mental workload (pp. 139-183). *Amsterdam: Elsevier.*

Khattak, Z.-H., Li, W., Karnowski, T., & Khattak, A-J. (2024). The role of driver head pose dynamics and instantaneous driving in safety critical events: Application of computer vision in naturalistic driving. *Accident Analysis and Prevention, 200 (2024)*: 107545.

Lee, J., Munoz, M., Fridman, L., Victor, T., Reimer, B., & Mehler, B. (2018). Investigating the correspondence between driver head position and glance location. *PeerJ Computer Science, 4(146):* 1-19.

Qin, L., Li, Z., Chen, Z. Andi Bill, M.S. & Noyce, D.A. (2019). Understanding driver distractions in fatal crashes: An exploratory empirical analysis. *Journal of Safety Research*, 69 (2019): 21-31.

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: Discriminating

styles of workload regulation and their safety impacts. *Journal of Experimental Applied Psychology, 19*(4): 287–300.

Schakmann, D., & Bosch, E. (2024). Automated Detection of Train Drivers' Head Movements: A Proof-of-Concept Study. *Automation, 5(1):* 35-48.

Zhang, L., Xiao, W., & Meng, D. (2019). Analysis of the correlation between Driver's Visual Features and Intention. *SAE,* Technical Paper 2019-01-1229.

Zhao, Y.; Görne, L., Yuen, L.M., & collaborators (2017). An orientation Sensor-Based Head Tracking System for Driver behaviour Monitoring. *Sensor, 17(11)* 2692*:* 1-16

Adaptive Automation of Human Information Processing in Driver-Road Interaction: A Driving Simulation Study in Motorway Environments

Alessandra Lioi [1*], Marco Bassani[1]

[1] Politecnico di Torino, Department of Environment, Land and Infrastructure Engineering, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy
[*]alessandra.lioi@polito.it

Abstract: This driving simulation study investigated how adaptive automation across the stages of human information processing (HIP) influences driving behaviour on motorways. Thirty-two participants navigated a motorway scenario under two traffic densities while interacting with an HUD in four interfaces displaying information according to the four stages of HIP, i.e., (i) Information Acquisition, (ii) Information Analysis, (iii) Decision Selection, and (iv) Action Implementation. These interfaces displayed key operational variables such as speed, lateral position, and distance to the preceding vehicle. A driving session without HUD was also included as a baseline. The findings show that adaptive automation of HIP levels influences driving behaviour, affecting speed regulation, longitudinal stability, and lateral control. These effects vary with traffic complexity and individual driver characteristics. Higher HIP levels of adaptive automation were linked to greater speed reductions under more demanding traffic conditions, indicating that interfaces offering advice and intervention become more relevant as traffic density increases. Changes in lateral positioning were also observed, though improvements were clearer for longitudinal control than lateral control. These results highlight the importance of evaluating not only traditional performance indicators but also system activation when assessing the effectiveness of in-vehicle interfaces. Additionally, the variability between individuals suggests that future systems could benefit from adaptive or personalised feedback strategies tailored to different driving behaviours.

1. Introduction

Driving involves a sequence of related stages typical of the so-called human information processing (HIP), i.e., (i) acquiring environmental cues, (ii) interpreting their meaning, (iii) selecting an appropriate response, and (iv) carrying out the necessary action. (Fig. 1, Parasuraman et al., 2000).



Fig. 1. *Human information processing framework*

Timely information processing is essential for safe vehicle control (Bubb et al., 2021), particularly at higher speeds, such as on motorways, where even minimal delays in perception or decision-making can increase safety risks. Drivers continuously regulate speed, maintain lane position, and control headway. These activities rely on constant visual monitoring and rapid motor responses. As a result, operational control is particularly vulnerable to distraction and attentional interference, especially on high-speed motorways.

The growing presence of in-vehicle technologies has introduced additional visual and cognitive demands that can interfere with the driver's ability to process critical roadway information promptly. Head-up displays (HUDs) have been proposed to reduce the need for drivers to divert their gaze from the roadway by presenting relevant information directly within the driver's field of view (Charissis et al., 2008). However, their effectiveness depends not only on where information is displayed, but also on how it is selected, organised, and presented. Poorly structured or excessively dense information may increase visual clutter and cognitive workload, potentially reducing situational awareness rather than improving it. The Adaptive Automation (AA) framework addresses this issue by selectively supporting different stages of the HIP while keeping the driver actively involved in vehicle control (Kaber et al., 2005). Within this framework, the HUD acts as the communication interface through which different levels of assistance are delivered.

Based on this framework, this driving simulation study examined how the adaptive automation of the different stages of the HIP affects driving performance and driver experience on motorways.

2. Method

2.1 Experimental design

The experiment considered a multi-level mixed factorial design. Thirty-two participants (age: M = 38.6 y, SD = 12.0 y; 16 females) drove in five configurations (HIP levels): a baseline without any support (level 0) and four HUD interfaces conveying information according to the adaptive automation of the four stages of HIP (levels 1 to 4). The road scenario consisted of a motorway, driven for about 8 km under two different traffic conditions, i.e., high and low traffic volumes (level of service, LOS C and LOS A).

2.2 Design of the HUD interfaces

The HUD interfaces were designed according to the HIP framework proposed by Parasuraman et al. (2000), which aimed to guide system designers in determining which functions of human information processing to automate and to what extent. The HUD displayed information on three key driving variables: speed, lateral position, and distance to the preceding vehicle (Fig. 2).

Automation can apply to input processes (data collection and analysis) and to output functions (decision-making, action selection, and execution).

At the Information Acquisition (1) level, the HUD provides real-time values. At Information Analysis (2), it processes the data and alerts the driver when necessary. At Decision Selection (3), it recommends the appropriate action, while at Action Implementation (4), the system executes the manoeuvre.

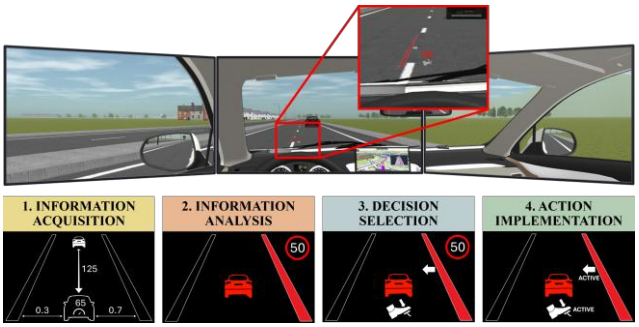


Fig. 2. *HUD interfaces*

2.3 Data analysis

Linear Mixed-Effects Models (LMMs) were used to analyse longitudinal and lateral driving behaviour, i.e., speed and lateral position (mean and standard deviation). The Participant ID was used as a random intercept.

3. Results

LMMs revealed significant effects of the automated HIP level, traffic conditions, and gender, as well as their interactions, on driving performance indicators. The random intercept for Participant ID was significant across all models, with Intraclass Correlation Coefficient (ICC) values demonstrating inter-individual variability. The difference between marginal and conditional R^2 further confirms that a considerable proportion of variance is attributable to random effects, supporting the use of mixed-effects models.

Regarding longitudinal control, mean speeds were lower, especially at the automated HIP levels 3 (Decision) and 4 (Implementation), compared with the baseline without HUD (Fig. 3a). In addition, the Implementation level produced significantly lower speeds than the Information level (HIP level 1). The level of service affects speed both as a single factor and in interaction with gender and automated HIP level. As expected, speeds were lower in heavier traffic conditions (LOS C, t_{278} = -2.051, p = .041) and, when considering the interaction with gender, post hoc tests revealed that males adopted a higher speed than females in LOS C (t_{37} = 2.653, p_{Holm} = .035). Furthermore, the reduction in speed associated with the automated HIP level 4 was particularly evident under higher traffic volumes (t_{278} = -4.497, p_{Holm} < .001).

The effect of the Implementation level in the longitudinal control was further confirmed by the analysis of speed variability. Speed variability was higher at automated HIP level 4 than at the baseline, Information, and Analysis levels (Fig. 3b), suggesting more frequent speed adjustments. Additionally, LOS C produced greater speed corrections than LOS A (t_{282} = 6.322, p_{Holm} < .001) and higher speed variation in females (t_{278} = -2.587, p_{Holm} = .039).

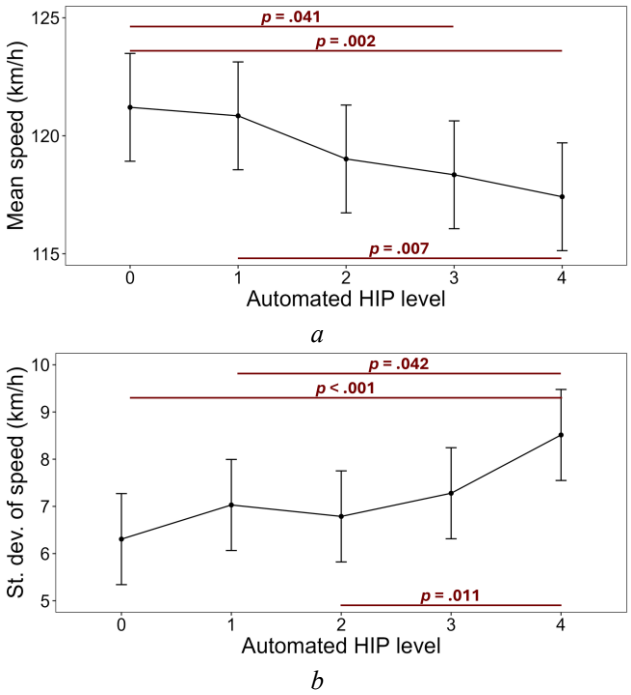


Fig. 3. *LMM results. Effect of HIP level on* ***(a)*** *mean speed and* ***(b)*** *standard deviation of speed*

Regarding lateral control, automated HIP levels significantly influenced the mean lateral position (LP), indicating systematic within-lane repositioning relative to the baseline across all HIP levels (Fig. 4a). Effects on lateral variability (SDLP) were less consistent. Only the automated HIP level 4 slightly improved lateral position stability relative to the baseline (Fig. 4b, t_{278} = -2.663, p_{Holm} = .082).

However, significant interaction terms (e.g., LOS * Gender; HIP * Gender) demonstrate that lateral stability was

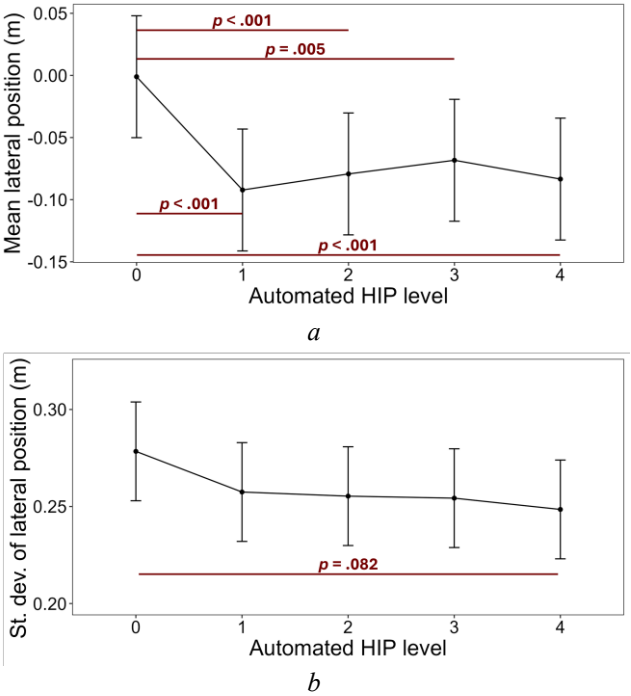


Fig. 4. *LMM results. Effect of HIP level on* ***(a)*** *mean lateral position and* ***(b)*** *standard deviation of lateral position*

sensitive to both environmental demands and individual characteristics. In particular, post hoc tests showed that male lateral stability benefited with Information (t_{278} = -3.493, p_{Holm} = .014) and Implementation (t_{278} = -3.081, p_{Holm} = .055) levels relative to the baseline scenario.

4. Discussion and conclusions

The results indicate that the adaptive automation of HIP influences speed regulation, longitudinal stability, and lateral control (Topliss et al., 2019). These effects appear to depend not only on the automated HIP level but also on traffic complexity and driver characteristics.

The speed reduction observed at higher HIP levels was more pronounced under LOS C traffic conditions, suggesting that the adaptive automation support becomes more relevant as traffic density increases. In more demanding traffic situations, drivers may benefit from timely advice (HIP level 3) and system interventions (HIP level 4) that support safer longitudinal behaviour. Regarding speed variability, the more ecological HIP level 3 led to fewer speed corrections, confirming Schewe and Vollrath (2020).

Regarding lateral behaviour, drivers tended to make slight adjustments to their lane position when interacting with the HUD at all HIP levels. Although these shifts were relatively small, they suggest that adaptive automation across HIP levels influenced drivers' spatial positioning strategies, regardless of HIP level. Processing visual information from the HUD may affect steering control and lane positioning.

At the same time, the limited improvements in lateral stability indicate that the benefits of higher HIP levels may be more significant for longitudinal control than for lateral control.

Finally, results confirm that drivers differ considerably in their responses to interface feedback and traffic conditions. Moreover, significant interaction effects involving gender suggest that individual characteristics influence how drivers operate across different automated HIP levels, posing challenges for developing effective driver support technologies (Park & Park, 2019).

In conclusion, the effect of adaptive automation in HIP on driving performance varied with traffic conditions and driver gender, particularly in speed regulation and lateral control. The benefit of adaptive automation in the Decision and Implementation levels, i.e., in suggesting and implementing driving manoeuvres, is demonstrated, especially for speed behaviour, by reducing the need for continuous driver adjustments and promoting more stable speed behaviour.

The observed variability between individuals indicates that drivers respond differently, suggesting that future systems may benefit from adaptive or personalised feedback strategies that account for individual driving styles.

Further research is needed to evaluate system activation exposure and to investigate these effects under real-world conditions, as the frequency and duration of system interventions may influence how drivers perceive and adapt to different levels of automation. Moreover, integrating additional measures of driver attention (e.g., gaze behaviour) and mental workload would provide a more comprehensive understanding of how drivers allocate their attention and cognitive resources when interacting with the HUD.

References

Bubb, H., Bengler, K., Grünen, R. E., & Vollrath, M. (2021). Automotive Ergonomics. Springer Nature.

Charissis, V., Papanastasiou, S., & Vlachos, G. (2008). Comparative Study of Prototype Automotive HUD vs. HDD: Collision Avoidance Simulation and Results (SAE Technical Paper 2008-01–0203). SAE International. https://doi.org/10.4271/2008-01-0203

Kaber, D. B., Wright, M. C., Prinzel III, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human factors*, *47*(4), 730-741. https://doi.org/10.1518/001872005775570989

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30(3), 286–297. https://doi.org/10.1109/3468.844354

Park, K., & Im, Y. (2020). Ergonomic guidelines of head-up display user interface during semi-automated driving. Electronics (Switzerland), 9(4). Scopus. https://doi.org/10.3390/electronics9040611

Park, J., & Park, W. (2019). Functional requirements of automotive head-up displays: A systematic review of literature from 1994 to present. Applied ergonomics, 76, 130-146. https://doi.org/10.1016/j.apergo.2018.12.017

Schewe, F., & Vollrath, M. (2020). Ecological interface design effectively reduces cognitive workload–The example of HMIs for speed control. Transportation research part F: traffic psychology and behaviour, 72, 155-170. https://doi.org/10.1016/j.trf.2020.05.009

Topliss, B. H., Pampel, S. M., Burnett, G., & Gabbard, J. L. (2019). Evaluating head-up displays across windshield locations. 244–253. https://doi.org/10.1145/3342197.3344524

Where Luminance Guides Driving: Route-Specific Attention Revealed by Vehicle Control and Pupil Responses

Jumpei Yamashita [1*], Yuta Suzuki [1], Fumika Nakanishi [2], Minori Yamataka [2], Hiroki Terashima [1], Kazushi Maruya [1]

[1] NTT Communication Science Laboratories, Atsugi, Japan
[2] DENSO CORPORATION, Kariya, Japan
[*] junpei.yamashita@ntt.com

Abstract:
Although safe driving is often assumed to rely on detailed foveal object recognition, drivers may also use coarse peripheral scene information, which gives a quick sense of the scene, to support control. Here we asked whether drivers use a coarse luminance grid of the forward scene—a low-resolution summary of how bright each region is and how it changes over time—during real-world driving. What matters is not the overall brightness but its spatial pattern across the scene. For 11 drivers on an urban route, we quantified, road segment by road segment, how strongly this luminance pattern predicted the driving-control score (steering, pedal, and speed) and how strongly the same pattern was reflected in the driver's pupil—a physiological signal of the luminance the visual system takes in. We then asked whether the two covaried across segments; they did, with segments where the pattern better predicted driving also showing greater pupil sensitivity (driver-level commonality, p = .01). This correlational evidence suggests that drivers may use coarse, driving-relevant luminance structure that the visual system registers. Practically, this points to a candidate marker of driver inattention: when driving-relevant luminance structure is present in the forward scene but the pupil does not reliably track it, a driver may be failing to take it in.

1. Introduction

Naturalistic driving requires rapid situation assessment and continuous control in a cluttered environment (B. Wolfe et al., 2022). Safe driving cannot depend on high-acuity foveal recognition of discrete objects alone (Summala et al., 1996; B. Wolfe et al., 2017). Scenes contain more than sequential fixations can process (J. M. Wolfe, 2021), and hazards must often be detected and evaded within a few hundred milliseconds (B. Wolfe et al., 2020; J. M. Wolfe et al., 2011). Drivers may therefore rely on faster, wide-field coarse processing, the extraction of scene structure well beyond the fovea within about 100–150 ms (Greene & Oliva, 2009; Thorpe et al., 1996). It can capture behaviorally relevant structure, such as dark regions marking nearby obstacles, even while gaze stays on the near road.

Here we asked whether drivers use a scene luminance grid, a highly reduced, panel-wise summary of the forward scene over time. Within this grid, overall brightness is a trivial cue; the informative signal is its spatial structure, the cell-to-cell deviations from overall brightness. How informative it is should vary along the route, weak on open road but richer in structured urban segments such as T-junctions (Oliva & Torralba, 2001; B. Wolfe et al., 2020). We then asked whether the same pattern was registered in the pupillary light reflex: the pupil responds to extrafoveal spatial structure as well as to overall luminance, and to whether that information is taken in (Binda & Murray, 2015; Binda et al., 2013). If the driving-predictive pattern also predicted the pupil, that match would be evidence that drivers use the grid.

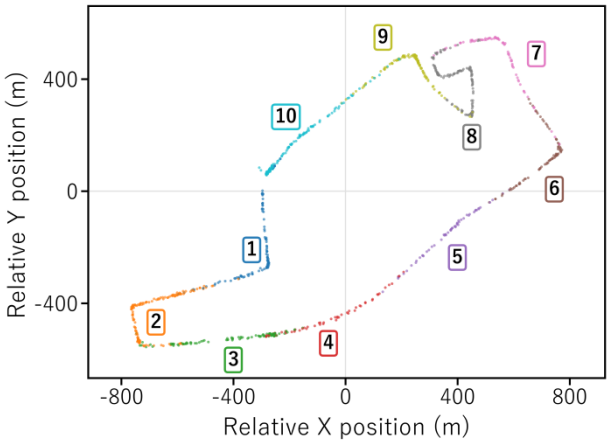


Fig. 1. *Urban driving route and the 10 analysis segments. Dots are all recorded vehicle GPS positions in relative X-Y coordinates (m), colored by segment; the numbers 1–10 label the segments in the order in which they are driven.*

To this end, we recorded the scene ahead, reduced the video to a coarse luminance grid, and modeled its relation to a driving-control score and pupil diameter across segments. Across drivers, the grid's spatial structure predicted driving beyond overall brightness, and the segments where it did so best were also those where the pupil was most sensitive to it. Such a match might point to a candidate marker of driver inattention, driving-relevant structure that the pupil does not follow.

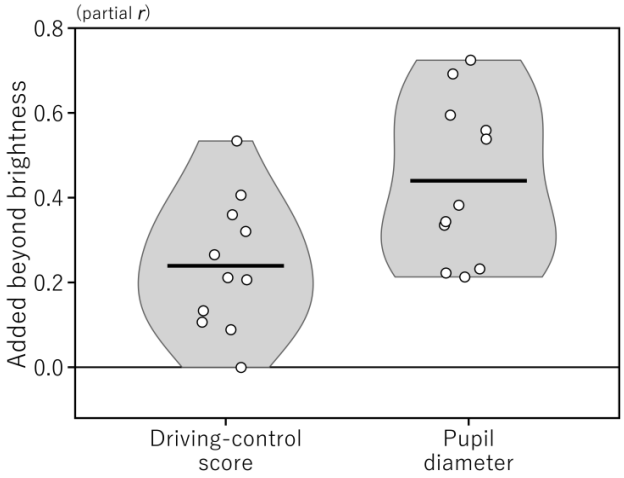


Fig. 2. *Contribution of spatial structure beyond overall brightness, shown as the partial correlation for each driver. Violins show the distribution over the 11 drivers, open circles individual drivers, and the bar the mean; zero means no contribution beyond overall brightness.*

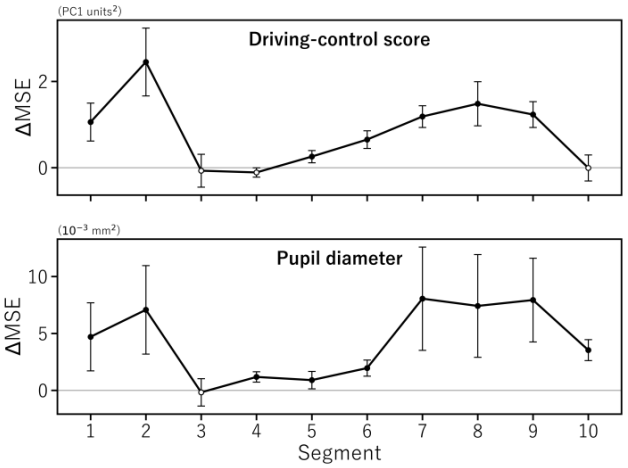


Fig. 3. *Contribution of spatial structure in each route segment, for the driving-control score (top) and pupil diameter (bottom). Means over 11 drivers, ±1 SEM; filled symbols mark positive values.*

2. Method

2.1 Data collection

Eleven drivers each drove about 45 min on a predefined urban route of about 10 km, two laps with 11 turns per lap. The route was divided into 10 equal GPS-based segments, each visited twice (Fig. 1). The forward scene was recorded with a centrally mounted RGB camera (1920 × 1200 pixels, 20 Hz), vehicle signals at 100 Hz, and pupil diameter with Tobii Glasses 3 at 100 Hz. Signals were aggregated into non-overlapping 20-s bins (974 in total, 80–96 per driver).

2.2 Scene luminance grid

Each scene-camera frame was converted to CIELAB and its L* channel, our proxy for scene luminance, reduced to a 6 × 8 grid of cell means; the top four rows (32 cells) were kept and the two bottom rows, showing only the instrument cluster and dashboard, discarded. Per-bin cell means were log-transformed and split into overall brightness (the mean log *L** across the 32 cells) and spatial structure (each cell's deviation from it).

2.3 Driving-control score and pupil measure

Six speed, steering, and pedal statistics were reduced to their first principal component (51% of the variance), computed once across all drivers, giving one score per bin. The score rose with busy control: steering movement, speed change, and accelerator use raised it, braking lowered it, and steady straight driving scored low. Pupil diameter was the bin mean after blink interpolation. The two are themselves almost uncorrelated (pooled r = .00 across bins; mean within-driver r = .08), so they are not, in effect, the same signal.

2.4 Base model and validation

For each driver, cross-validated ridge regression predicted both signals from the 32 log-luminance cells; every analysis below rests on these models. Models were fitted within driver, never across drivers, and cross-validated over the ten segments: each was held out in turn and predicted from a model fitted on the other nine. Both visits to a segment were held out together, so the model was never tested on a segment it had already seen in training, and every accuracy and importance reported below is out-of-fold.

3. Results

3.1 Spatial structure contributed beyond overall brightness

We first asked what the spatial structure adds beyond overall brightness alone, by partialing the brightness-only prediction out of the full-grid prediction (Fig. 2). Spatial structure contributed beyond brightness for both the driving-control score (partial r = .23) and pupil diameter (partial r = .42), in every driver (sign test p = .001).

3.2 Driving and pupil relied on the same structure

Within each segment we scrambled the spatial structure across time bins, leaving overall brightness untouched, and took the loss of accuracy on re-prediction as its contribution there. This gave, for each driver, a driving profile and a pupil profile: ten numbers each, showing how much spatial structure mattered in each segment (Fig. 3). We then removed from each profile what the segment's own brightness and variability could explain.

We then split the 11 drivers into groups of five and six, averaged the profiles within each group, and correlated the driving profile of one group with the pupil profile of the other across the ten segments. Because the profiles came from different people, nothing shared within a driver could create it. We repeated this over all 924 splits and averaged.

The remaining correlation was positive (mean partial r = .39; p < .01 by permutation; 95% CI [+.05, +.74], obtained by recomputing the analysis with each driver left out in turn) (Fig. 4).

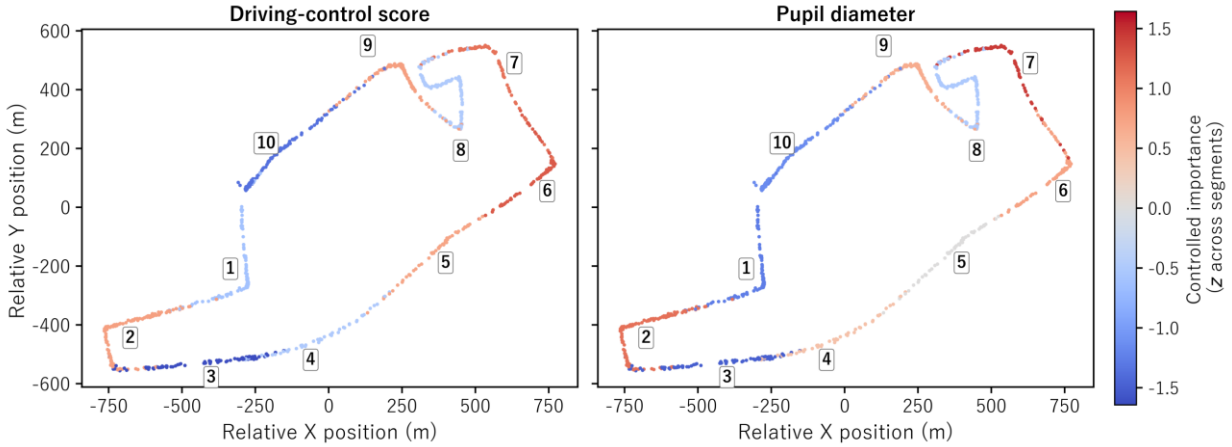


Fig. 4. Contribution of spatial structure mapped onto the route, after removing what each segment's own brightness and variability could explain, for the driving-control score (left) and pupil diameter (right).

4. Discussion

Drivers appear to use coarse luminance structure during naturalistic control, at least on the indirect evidence of a correspondence between its contributions to driving and to the pupil. Coarse processing offers a fast alternative, and should pay off where luminance covaries with road geometry and obstacles (Oliva & Torralba, 2001). For inattention research, this points to a marker that needs only the forward scene and pupil size, not a measurement of where the driver is looking.

What links the two is not settled by our data. One candidate is attention: attention modulates the pupillary light reflex beyond its raw response to luminance (Binda & Murray, 2015; Binda et al., 2013), so drivers might weight behaviorally relevant structure where it is informative.

The coupling did not depend on how the route was cut: we shifted the segment boundaries by a quarter, a half, and three quarters and averaged the four results, and it stayed positive and above its permutation null ($p < .05$) at every granularity from six to fifteen, though it weakened as segments grew shorter and their estimates noisier. Each profile is ten values saying how much the luminance pattern mattered per segment. The $r = .39$ above is what remained after predicting these from the segment's own brightness and variability. Removing the segment's average speed and accelerator use, the driving variables most related to where the coupling appeared, still left $r = .31$ ($p < .01$). The permutation null was itself centered near zero ($r = .01$), so the procedure cannot manufacture a positive value. The estimates are nonetheless noisy. Each segment contributes only about nine observations per driver, and cutting the route finer would only thin them further, so the correlation rests on just ten segments.

5. Conclusions

A highly reduced luminance grid predicted the driving-control score, and the same structure was registered, more weakly, by the pupil in the same segments. The forward scene can therefore set an expectation for the pupil, and a pupil that falls short of it may mark a driver who has not taken the structure in. Experimental manipulation of luminance and attentional demand is needed to test this.

References

Binda, P., & Murray, S. O. (2015). Spatial attention increases the pupillary response to light changes. *Journal of Vision, 15*(2), 1. https://doi.org/10.1167/15.2.1

Binda, P., Pereverzeva, M., & Murray, S. O. (2013). Attention to bright surfaces enhances the pupillary light reflex. *Journal of Neuroscience, 33*(5), 2199–2204. https://doi.org/10.1523/JNEUROSCI.3440-12.2013

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20*(4), 464–472. https://doi.org/10.1111/j.1467-9280.2009.02316.x

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175. https://doi.org/10.1023/A:1011139631724

Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors, 38*(3), 442–451. https://doi.org/10.1518/001872096778701944

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522. https://doi.org/10.1038/381520a0

Wolfe, B., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the Useful Field: Considering peripheral vision in driving. *Applied Ergonomics, 65*, 316–325. https://doi.org/10.1016/j.apergo.2017.07.009

Wolfe, B., Sawyer, B. D., & Rosenholtz, R. (2022). Toward a theory of visual information acquisition in driving. *Human Factors, 64*(4), 694–713. https://doi.org/10.1177/0018720820939693

Wolfe, B., Seppelt, B., Mehler, B., Reimer, B., & Rosenholtz, R. (2020). Rapid holistic perception and evasion of road hazards. *Journal of Experimental Psychology: General, 149*(3), 490 - 500. https://doi.org/10.1037/xge0000665

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review, 28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences, 15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Driver Distraction and Mobile Device Usage: Safety Performance Indicator Results for Croatia

Juraj Leonard Vertlberg [1*], David Gruonjić [1], Slaviša Babić [2], Marijan Jakovljević [1]

[1] University of Zagreb Faculty of Transport and Traffic Sciences, Vukelićeva 4, Zagreb, Croatia
[2] Croatian Roads, Ltd., Vončinina 3, Zagreb, Croatia
[*] jvertlberg@fpz.unizg.hr

Abstract: Driver distraction represents a significant behavioural risk factor affecting road safety. Among different sources of distraction, the use of mobile devices while driving has become increasingly common and is associated with reduced driver attention, longer reaction times and a higher probability of traffic crashes. Within the European road safety monitoring framework, driver distraction is measured through Key Performance Indicators (KPIs), enabling the assessment and comparison of safety-related behaviours across countries. The objective of this research is to estimate the national value of the KPI for driver distraction in Croatia, defined as the share of drivers that do not use a mobile device while driving. Data were collected through roadside observational surveys conducted under free-flow traffic conditions across the national road network. Observations were carried out at 41 locations covering motorways, rural and urban roads, resulting in a total sample of 22,675 driver observations. The results show that approximately 94% of drivers in Croatia do not use mobile devices while driving. The highest compliance was observed on rural roads (96%), while motorways and urban roads recorded compliance levels of around 94%. Although most drivers demonstrate safe behaviour, mobile device usege while driving remains a relevant road safety concern.

1. Introduction

Driver distraction is recognised as an important behavioural risk factor affecting road safety. It occurs when a driver's attention is diverted from activities necessary for safe driving towards competing tasks. Among various forms of driver distraction, the use of mobile devices while driving has become increasingly common and has been shown to impair driving performance, increasing reaction times and reducing driver attention as well as vehicle control (Lipovac et al., 2017). Activities such as texting or interacting with smartphone applications introduce visual, manual and cognitive distraction simultaneously, significantly increasing crash risk (Oviedo-Trespalacios et al., 2016; Guo & Lu, 2022; NHTSA, 2024).

In Croatia, mobile device usage has contributed to 12% of severe traffic accidents, as shown in Figure 1 (Šoštarić et al., 2025). Despite a general trend of decreased number of road traffic accidents, the number of mobile device usage related accidents presents a key concern, both in Croatia and in other countries.

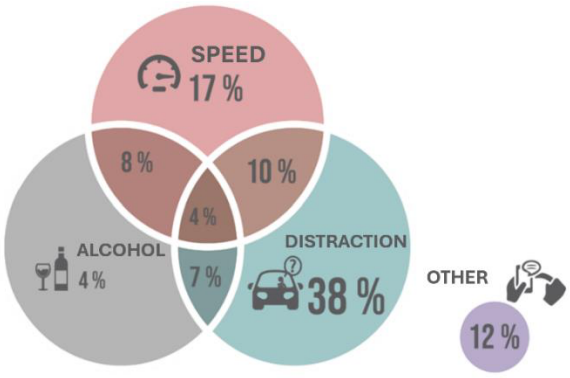


Fig. 1. *Key factors contributing to severe traffic accidents in Croatia*

To support evidence-based road safety policies, the European Union promotes the use of Key Performance Indicators (KPIs) that allow monitoring of behavioural risk factors within the road transport system. Within this framework, driver distraction is measured as the share of drivers who do not use mobile devices while driving.

The objective of this study is to estimate the national value of the driver distraction KPI in Croatia based on observational roadside data collected across different road types and vehicle categories.

2. Method

2.1 KPI Definition

In accordance with the methodological guidelines for road safety indicators, the KPI for driver distraction is defined as the share of drivers that do not use a mobile device while driving (European Commission, 2023). Mobile device use while driving reduces the driver's level of attention, shifts focus and increases driver's reaction time, thereby potentially increasing the risk of road traffic crashes occurence.

The indicator is therefore calculated as the share of observed drivers who were not using a mobile device while driving.

2.2 Data Collection Procedure

Data were collected through roadside observational surveys conducted under free-flow traffic conditions. Observations were performed at road cross-sections where vehicles were travelling without interruption. Vehicles that were stationary or temporarily stopped, for example at traffic lights or in traffic queues, were excluded from the analysis.

During each data collection session, traffic counts were conducted simultaneously in order to enable appropriate weighting of the results and estimation of statistical reliability. Observations were carried out during daylight hours and

under favourable weather conditions in order to ensure consistent observation conditions.

2.3 Sampling Design and Study Locations

To ensure national representativeness and methodological comparability, data collection followed predefined sampling requirements. Observations were conducted across three road categories: motorways, rural roads outside urban areas, and urban roads within built-up areas. A minimum of 500 observations per road category and at least ten locations per road type were required.

In total, data were collected at 41 locations across Croatia, including 18 urban, 13 rural and 10 motorway sites. Observations were conducted on both weekdays and weekends, resulting in 61 observation sessions. Locations were randomly selected and geographically distributed to represent different parts of the national road network (Faculty of Transport and Traffic Sciences, 2025). Figure 2 shows the observation locations by road category.



Fig. 2. *Locations of roadside observations in Croatia*

2.4 Sample Characteristics

Across all observation locations, a total of 22,675 driver observations were recorded. The sample included the most common vehicle categories present in road traffic:

- 19,869 passenger cars
- 1,197 light goods vehicles
- 1,228 heavy goods vehicles
- 381 buses

The large sample size and geographical distribution of observation sites ensured that the collected dataset provides a reliable basis for estimating the national value of the distraction KPI.

3. Results

The collected observational data were analysed to determine the share of drivers who were not using mobile devices while driving. In accordance with the methodological requirements, the primary KPI value was calculated using observations collected during weekdays under daytime conditions.

3.1 Results by Road Type

The analysis indicates high compliance levels across all road categories. On motorways, approximately 94% of drivers were observed not using mobile devices while driving. Rural roads showed slightly higher compliance, with approximately 96% of drivers not using mobile devices. Urban roads recorded a compliance level of approximately 94%. Figure 3 shows the compliance levels per road category.

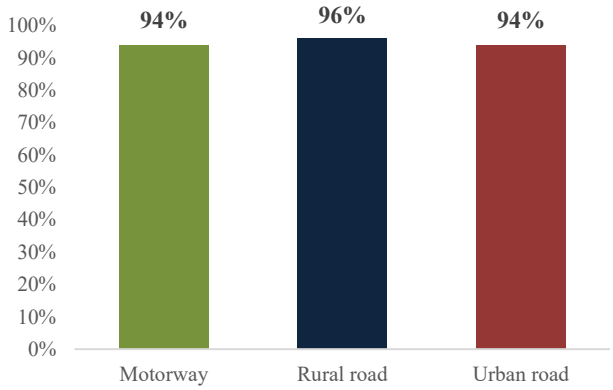


Fig. 3. *Compliance levels per road category*

When considering all road categories combined, the overall KPI value indicates that approximately 94% of drivers do not use mobile devices while driving. These results suggest that although the majority of drivers comply with safe driving practices, a non-negligible share of drivers continue to engage in distracting activities while operating a vehicle.

3.2 Results by Vehicle Category

Differences were observed between vehicle categories. Passenger car drivers showed compliance levels of approximately 94–95%, consistent with the national average, while light goods vehicle drivers demonstrated slightly lower compliance (about 91%). The lowest compliance was observed among heavy goods vehicle drivers, with approximately 89% not using mobile devices while driving. Bus drivers showed relatively high compliance levels (around 94%). Figure 4 presents compliance levels by vehicle category across all locations.

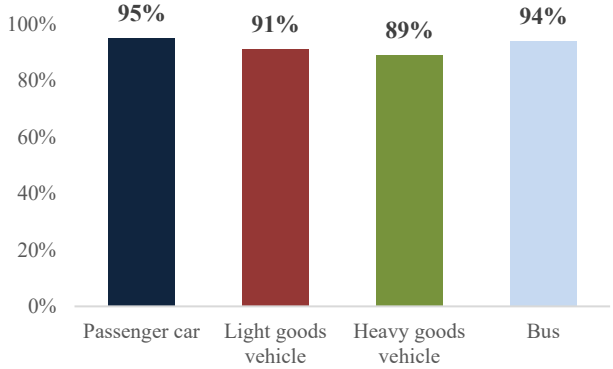


Fig. 4. *Compliance levels per vehicle category*

These findings suggest that professional drivers, particularly those operating heavy goods vehicles, may be more likely to engage in mobile device use while driving.

3.3 Results by Observation Period

Additional analysis examined differences between weekday and weekend observations. The results indicate that compliance levels were slightly higher during weekend observations, reaching approximately 97% of drivers not using mobile devices. Figure 5 shows compliance levels during working days and weekends on all locations.

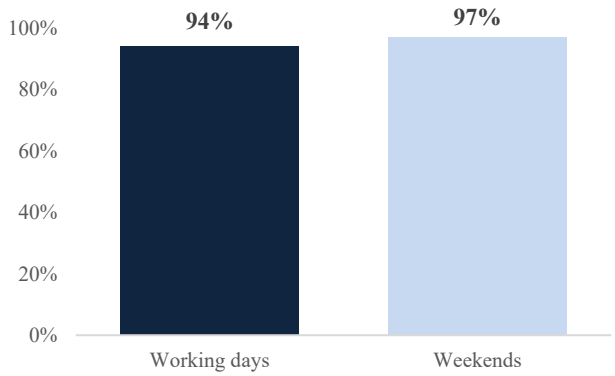


Fig. 5. *Compliance levels on working days and weekends*

When considering all observation periods combined, the overall KPI value across the entire dataset was approximately 95%.

4. Discussion

The results indicate relatively high compliance with safe driving behaviour in Croatia regarding mobile device use while driving. The overall KPI value of approximately 94% suggests that most drivers refrain from using mobile devices while operating a vehicle, although the remaining share still represents a road safety concern.

Only small differences were observed between road types. Slightly higher compliance on rural roads compared with motorways and urban roads may be related to differences in traffic density, driving conditions or perceived enforcement levels. Urban environments, with more frequent stops and interactions with traffic control devices, may provide greater opportunities for drivers to use mobile devices.

Differences were also observed between vehicle categories. Lower compliance among heavy goods vehicle drivers may reflect the communication demands associated with professional driving, such as coordinating logistics and delivery activities. Nevertheless, distraction among professional drivers can pose significant risks due to the size and mass of the vehicles involved.

Although the share of drivers using mobile devices appears relatively small, previous research shows that even occasional distraction can increase crash risk. Continuous monitoring of driver distraction indicators is therefore important for identifying behavioural trends and supporting targeted road safety interventions.

5. Conclusions

The results indicate relatively high compliance with safe driving behaviour in Croatia, as the majority of drivers were observed not using mobile devices while driving. Nevertheless, the share of drivers who continue to use mobile devices while operating a vehicle still represents a road safety concern.

Differences between road types were relatively small, although slightly higher compliance was observed on rural roads compared with motorways and urban roads. Lower compliance among heavy goods vehicle drivers may reflect the operational demands placed on professional drivers who rely more frequently on communication during travel.

Although the observed prevalence of mobile device use while driving appears limited, even small percentages of distracted drivers can significantly increase crash risk. Continuous monitoring of driver distraction indicators therefore remains important for evaluating behavioural trends and supporting the development of effective road safety measures.

References

European Commission. (2023). Methodological guidelines for the calculation of key performance indicators: Distraction. Publications Office of the European Union.

Guo, F., & Lu, D. (2022). How many crashes does cellphone use contribute to? Population attributable risk of cellphone use while driving. Journal of Safety Research, 82, 385–391.

Lipovac, K., Đerić, M., Tešić, M., Andrić, Z., & Marić, B. (2017). Mobile phone use while driving: Literature review. Transportation Research Part F: Traffic Psychology and Behaviour, 47, 132–142.

Faculty of Transport and Traffic Sciences, University of Zagreb. (2024). TRENDLINE project report: Road safety key performance indicators in Croatia.

National Highway Traffic Safety Administration. (2024). Distracted driving. https://www.nhtsa.gov/risky-driving/distracted-driving

Regan, M. A., Lee, J. D., & Young, K. L. (Eds.). (2009). Driver distraction: Theory, effects, and mitigation. CRC Press.

Šoštarić, M., Ševrović, M., Jakovljević, M., & Gruhonjić, D. (2025). Analysis of critical factors influencing the occurrence of traffic accidents in Croatia. Transportation Research Procedia, 83, 472–479.

Extended Abstract for the 10th International Conference on Driver Distraction and Inattention (DDI2026)

Enhancing driver attention through behavior change interventions: Insights from the Norwegian driving challenge

Anine Riege [1], Irmelin Bergh [1], Rita Helen Aarvold [2*]

[1] Nudgelab, Oslo, Norway
[2] The Norwegian Public Roads Administration, Traffic Safety, Transport and Society, Oslo, Norway
[*] rita.aarvold@vegvesen.no

Abstract: The Norwegian Public Roads Administration conducted the behavioral intervention *The National Driving Challenge* as part of the national driver distraction campaign "Thank you for your attention". The intervention was developed to address the intention–action gap identified in a national baseline survey and was informed by the Health Action Process Approach (HAPA), implementation intentions, and self-regulation theory. Following a controlled pilot study, the intervention was refined before national implementation. Over an eight-week period, participants (N=896) selected one distracted behavior to discontinue from a predefined list in order to increase their attention as drivers. The intervention incorporated techniques such as self-monitoring, implementation intentions, and reflective logging. The participants received weekly motivational text messages (nudges). Eighty-nine percent reported successful behavior change, and 70% showed spillover to additional behaviors. Planning and self-efficacy were the strongest motivational-phase predictors of change, and self-monitoring the strongest in the action-control phase, whereas attitudes toward attentive driving remained unchanged. The findings indicate that targeted behavior-change strategies can strengthen driver attention and that future road safety campaigns and related initiatives should further explore behavior-focused approaches.

1. Introduction

Driver distraction is a major road-safety problem in Norway and internationally. In-depth analyses of fatal crashes on Norwegian roads indicate that driver inattention or distraction contributes to close to one in three cases (Aasvik & Sagberg, 2023). Because attention is a prerequisite for safe driving, reducing distraction is central to lowering crash risk. Among the available countermeasures, influencing drivers' behavior and attitudes has long been a core element of the Norwegian Public Roads Administration's (NPRA) safety work, which treats sustained, multi-component campaigns as an important tool for promoting safe behavior.

As part of this work, NPRA launched the national driver-distraction campaign "Takk for oppmerksomheten" ("Thank You for Your Attention") in 2018, with a relaunch in 2023. A baseline survey of 3,025 Norwegian drivers revealed a clear intention–action gap: most drivers intend to drive attentively but struggle to translate that intention into behavior (Opinion, 2023). This gap is well captured by the Health Action Process Approach (HAPA; Schwarzer, 2008), which distinguishes a motivational phase, in which intentions form, from a volitional phase, in which intentions are enacted, and identifies self-efficacy, planning, and action control as the bridges between the two.

Distracted behaviors such as eating, checking a phone, or daydreaming are often automatic and habitual (Kahneman, 2011; Lally & Gardner, 2013), which makes them resistant to attitude-based appeals but responsive to volitional, action-oriented techniques. Accordingly, the intervention combined implementation intentions specifying when and how to act (Gollwitzer, 1999), daily self-monitoring, and weekly motivational nudges (Thaler & Sunstein, 2008) timed to the fresh-start effect (Dai, Milkman, & Riis, 2014). Focusing effort on a single, concrete behavior also draws on goal-setting principles (Locke & Latham, 1990) and may generate behavioral spillover to related behaviors (Dolan & Galizzi, 2015).

Building on the campaign evaluation and the 2023 baseline, NPRA developed "The Driving Challenge," an eight-week, behavior-focused intervention in which drivers voluntarily chose one distracted behavior to discontinue (Bergh & Riege, 2025). The initiative was first evaluated in a controlled pilot study including a test and comparison group. Findings from the pilot informed the refinement and subsequent national implementation of the intervention. The present study examines whether this targeted, behavior-focused approach can strengthen driver attention, and which self-regulatory factors predict successful change.

2. Method

2.1 Design

The National Driving Challenge was a voluntary, single-arm pre–post intervention delivered to a self-selected national sample.

2.2 Participants

Recruitment was carried out by the research agency Opinion. Eligible participants drove daily or several times a week and selected one specific distracted behavior they

wished to change. Of 4,075 registrants, 2,554 completed the eight-week challenge, and 896 completed both the pre- and post-intervention surveys and all eight weeks of behavior logging; this complete-case group formed the analytic sample. The sample was nationwide and predominantly male (67.6%), with all age groups represented (median age group 50–59 years).

2.3 Procedure

From a predefined list, participants selected one distracted behavior to discontinue (e.g., eating, daydreaming, checking a phone). Each participant then formed an implementation intention specifying when and how they would change the behavior, and kept daily logs as a form of self-monitoring over the eight weeks. Every Monday, participants received a motivational text message (a nudge) timed to coincide with the fresh-start effect. The post-intervention survey was distributed in week 50, with two reminders.

2.4 Measures and analysis

Pre- and post-intervention surveys assessed self-reported attention while driving, perceived attention of other drivers, attitudes toward attentive driving, and the HAPA constructs of intention, self-efficacy, perceived control, and planning. Participants also submitted daily behavior logs.

3. Results

3.1 Behavior change

Overall, 89% of participants reported successfully changing their chosen behavior (a score of 7–10 on a 10-point self-report scale), while only 4.2% reported never or rarely succeeding (scores 1–4). Self-rated success was consistently high across all target behaviors, ranging from 7.8/10 for daydreaming to 9.0/10 for handheld phone use. Reported change was modestly greater among older participants, $r = .211$, $p < .001$, and unrelated to gender.

3.2 Spillover

Although participants selected only one behavior to discontinue, 70% reported changing at least one additional distracted behavior (40.6% changed one, 16.3% two, and the remainder three or more), indicating a substantial spillover effect.

3.3 Predictors of change

Two regression models tested HAPA constructs as predictors of behavior change. Motivational-phase factors jointly explained 17.0% of the variance, $F(4, 875) = 44.96$, $p < .001$, $R^2 = .170$, with planning ($\beta = .24$, $p < .001$) and self-efficacy ($\beta = .21$, $p < .001$) the strongest predictors; perceived control was weaker but significant ($\beta = .09$, $p = .006$), and intention was not ($\beta = .06$, $p = .074$). Action-control factors explained 17.8% of the variance, $F(3, 885) = 64.08$, $p < .001$, $R^2 = .178$, with self-monitoring the strongest predictor ($\beta = .24$, $p < .001$), followed by awareness of goals ($\beta = .17$, $p < .001$); effort was not significant ($\beta = .07$, $p = .091$).

3.4 Attention and attitudes

Self-regulation measures increased from pre- to post-intervention, though effect sizes were small. Participants' self-perceived inattention decreased while their perception of others' inattention rose slightly, $\Lambda = .95$, $F(1, 895) = 46.99$, $p < .001$, $\eta p^2 = .05$; throughout, participants judged others as far more inattentive than themselves, $\Lambda = .38$, $F(1, 895) = 1461.94$, $p < .001$, $\eta p^2 = .62$. Attitudes toward attentive driving did not change over the eight weeks and did not differ by age or gender.

4. Discussion

The National Driving Challenge produced widespread self-reported behavior change and improved self-perceived attention, yet left attitudes unchanged. This pattern is the study's central contribution: it suggests that asking drivers to discontinue one concrete, self-chosen behavior is a more effective route to attentive driving than appeals aimed at attitudes. The intervention operated on the volitional side of the intention–action gap rather than the motivational side, which is where earlier campaign work suggested drivers were falling short.

The predictors of change clarify the mechanism and support the HAPA framework (Schwarzer, 2008). Planning and self-efficacy predicted successful change during the motivational phase, while self-monitoring was the strongest action-control predictor, consistent with accounts of how repeated, monitored action supports new habits (Lally & Gardner, 2013). Notably, intention did not predict change — an empirical echo of the intention–action gap itself: intending to drive attentively was not enough, whereas planning, self-efficacy, and self-monitoring were. Taken together with the behavior-focused design, this suggests that campaigns should focus on changing one specific behavior, ideally supported by strategies that strengthen planning, self-efficacy, and self-monitoring.

Two further findings carry practical weight. The spillover to non-targeted behaviors (Dolan & Galizzi, 2015) suggests that a narrow, single-behavior focus can yield broader behavior change. The persistent, strengthening belief that other drivers are more inattentive than oneself is consistent with overconfidence effects (Kruger & Dunning, 1999) and helps explain the unchanged attitudes (Maio et al., 2019): most drivers already think they are attentive, so generic "be attentive" messaging has little impact, whereas a concrete behavioral target sidesteps this belief.

The study does, however, have limitations: All outcomes were self-reported, with no objective measure of driving behavior; the self-selected, highly motivated sample and substantial attrition (partly due to a delayed second survey) limit generalizability. Attrition analyses indicated only minor demographic and behavioral differences between completers and non-completers. The eight-week window is too short to establish durable change, and the correlational design precludes causal claims. Future work should incorporate objective or in-vehicle behavioral measures, a controlled design with longer follow-up, and tests of whether the spillover and attention gains persist. Even so, effect sizes that are modest per driver may translate into meaningful road-safety gains at population scale.

5. Conclusions

This nationwide intervention showed that most drivers were able to change a self-selected distracted behavior — with change often spilling over to other behaviors — even as their general attitudes remained unchanged. Targeting one specific behavior, rather than attitudes toward attentive driving, therefore appears to be a promising approach for reducing driver distraction. Road safety initiatives pursuing Vision Zero should further develop and rigorously test behavior-focused interventions of this kind.

References

Aasvik, O., & Sagberg, F. (2023). *Evaluering av den nasjonale oppmerksomhetskampanjen «Takk for oppmerksomheten»: En undersøkelse med fire ulike datainnsamlingsmetoder i perioden 2018–2022* (TØI-rapport 1949/2023). Transportøkonomisk institutt. https://www.toi.no/publikasjoner/evaluering-av-den-nasjonale-oppmerksomhetskampanjen

Dai, H., Milkman, K. L., & Riis, J. (2014). The fresh start effect: Temporal landmarks motivate aspirational behavior. *Management Science, 60*(10), 2563–2582. https://doi.org/10.1287/mnsc.2014.1901

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology, 47*, 1–16. https://doi.org/10.1016/j.joep.2014.12.003

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54*(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lally, P., & Gardner, B. (2013). Promoting habit formation. *Health Psychology Review, 7*(sup1), S137–S158. https://doi.org/10.1080/17437199.2011.603640

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice Hall.

Maio, G. R., Haddock, G., & Verplanken, B. (2019). *The psychology of attitudes and attitude change* (3rd ed.). SAGE.

Opinion. (2023). *Oppmerksomhet i trafikken: Referansemåling, mai–juni 2023* [Internal report prepared for the Norwegian Public Roads Administration]. Opinion AS.

Bergh, I., & Riege, A. (2025). *Den nasjonale kjøreutfordringen: Hvordan bli en mer oppmerksom sjåfør* [Unpublished report prepared for the Norwegian Public Roads Administration]. Nudgelab.

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology, 57*(1), 1–29. https://doi.org/10.1111/j.1464-0597.2007.00325.x

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Partial driving automation can impose additional demand on the driver without impairing driving performance in a car-following scenario

Katharina Kummerer [1,2*], Sebastian Hergeth [2], Christian Gold [2], Josef F. Krems [1]

[1] Cognitive and Engineering Psychology, Chemnitz University of Technology, Chemnitz, Germany
[2] BMW Group, Munich, Germany
[*] katharina.kummerer@bmw.de

Abstract: In a driving simulator study (N = 60), self-reported demand imposed on the driver by the driving task and secondary tasks, and their effect on driving performance were examined during manual and partially automated driving (L2). Reported demand was higher during L2 driving, particularly for system-paced tasks, whereas brake response times did not show a corresponding impairment. Results indicate that higher demands under L2 do not necessarily degrade driving performance surrogates, such as brake response time. Given that driver distraction is often assessed using gaze-based measures only, the current findings indicate that comprehensive test methods to evaluate 'sufficient' attention during secondary task engagement should also include driving performance metrics.

1. Introduction

A major cause of crashes is driver diverted attention to activities not related to driving, known as *driver distraction* (Regan et al., 2011). Distraction guidelines (e.g., National Highway Traffic Safety Administration [NHTSA], 2013) provide methods to evaluate distraction related to secondary tasks and principles for designing in-vehicle infotainment systems (IVIS) to support safe driving. These principles include that functions should be driver-paced, interruptible, and not exceed distraction levels of an accepted reference task (NHTSA, 2013). Test methods for distraction typically focus on off-road glances, while adaptations in driving behaviour are often overlooked and only implicitly inferred by linking off-road glances with increased crash risk (e.g., Klauer et al., 2006). Beyond established methods, an approach assessing 'sufficient' attention while driving could consider the allocation of spare capacity according to situational demands (e.g., Kircher & Ahlstrom, 2017; Kujala & Grahn, 2022). Additionally, incorporating behavioural adaptation mechanisms could capture the effects of self-regulation (Kummerer et al., 2025, 2026), thereby enhancing existing test methods.

Today's distraction guidelines were developed before widespread driving automation and advanced IVIS (e.g., NHTSA, 2013), and mainly address manual driving without assistance systems (L0; Society of Automotive Engineers [SAE], 2021). Research shows that partial driving automation (L2; SAE, 2021) can reduce demands by providing lateral and longitudinal driving support (Eddine et al., 2024). However, supervising system functions is mandatory (SAE, 2021) and increased off-road attention during L2 driving (Biondi & Jajo, 2024; Eddine et al., 2024) may increase the risk of missing system failures.

This research aims to contribute to the development of a test method for evaluating 'sufficient' attention during distraction in L0 and L2 driving by integrating behavioural measures to assess when driver capacities are exceeded and safe driving is compromised. Derived from the Driver Workload Metrics model (Fig. 1; Angell et al., 2006; Hurts et al., 2011), the following hypotheses were tested:

H1: The use of an L2 system reduces self-reported demand on the driver (hereafter referred to as *demand*).

H2: Secondary tasks increase demand, especially when not being driver-paced.

H3: Increased demand leads to impairments in driving performance, such as brake response time.

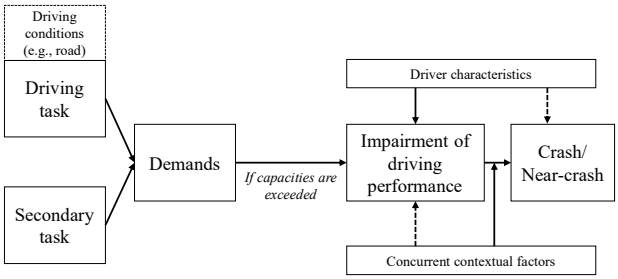


Fig. 1. *Driver Workload Metrics model [adapted and modified from Angell et al. (2006, p. xxviii) and Hurts et al. (2011, p.7)]*

2. Method

2.1 Participants

A driving simulator study was conducted with N = 60 participants (23 women). Average age was 36.72 years (SD = 8.93). All participants had held a valid driving license for an average of 19.12 years (SD = 8.69).

2.2 Apparatus

The static driving simulator featured a 220° horizontal field of view using a LED wall (7200×1800 pixels), with side and rear mirrors shown on displays. The secondary task was displayed on the center information display (CID; 2880×960 pixels). The simulator ran on SPIDER software (Strobl, 2003).

2.3 Study design and procedure

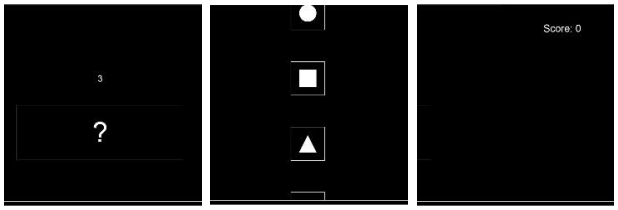
A mixed design examined the effects of driving mode (between) and secondary tasks (within) on self-reported demand and their relationship with brake response time.

Participants, randomly assigned to drive at L0 or L2 (L2: $n = 31$), performed self-paced and system-paced visual-manual tasks and a reading task on the CID during randomised experimental drives. A baseline drive without secondary tasks served as reference. Secondary tasks were presented at predefined intervals. Unlike driving at L0, system failures requiring participants to supervise the automation were introduced in the L2 group. Participants were instructed to prioritise driving and engage in secondary tasks only when conditions allowed. After each drive, participants completed the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) to assess demand from both the driving task and the secondary task.

In a car-following scenario, sudden braking by the lead vehicle soon after secondary task onset, combined with system failure while driving L2, increased rear-end collision risk, a crash type often linked to distracted driving (e.g., Eby & Kostyniuk, 2003; Najm & Smith, 2007; Sundfør et al., 2019).

2.4 Secondary tasks

Visual-manual tasks were adapted from Liang et al. (2024). Participants searched for a target (triangle) among distractors either by scrolling through a list of geometric elements at their own pace (self-paced) or by searching within a continuously moving list (system-paced). Self-paced tasks began with a 4-s time window in which the task could be accepted or declined (Metz et al., 2011) (Fig. 2). System-paced tasks were announced by a 4-s timer without the option to decline. Participants earned 10 points per identified triangle and lost 20 points for each severe driving error, in order to incentivise secondary task engagement while simultaneously emphasising the primary importance of the driving task. The reading task text length matched typical check-control messages shown during driving.



(a) (b) (c)

Fig. 2. *Self-paced visual-manual task (a) Choice, (b) Task, (c) Score*

2.5 Data analysis

The evaluation of interdependencies between demand and driving performance focused on the sudden braking event. Demand was measured using the NASA-TLX (Hart & Staveland, 1988) with an added visual demand item, each rated on a 20-point scale. Brake response time to the lead vehicle's braking was used as a surrogate measure for driving performance. Influencing factors were analysed using Linear Mixed Models (LMMs) with participant as a random effect. Due to assumption violations, confidence intervals were estimated via bootstrapping (n = 5,000; Leeuw & Meijer, 2008). Interaction terms were not included, as they did not improve model fit.

3. Results

3.1 Influences on self-reported demand

NASA-TLX scores indicated (Fig. 3) that self-reported demand was highest for the system-paced condition in both L0 ($M = 58.3$, $SD = 23.6$, $n = 22$) and L2 ($M = 72.4$, $SD = 23.2$, $n = 27$).

Analyses (see Appendix A, Table 1) revealed significant positive associations between all fixed effects and demand ($p < .01$). Therefore, H1 could not be confirmed, while H2 was supported. The lack of model improvement from interaction terms indicates that the relationship between secondary tasks and demand was independent of driving at L0 or L2. The variance indicators revealed high variance between participants ($\sigma^2 = 245.0$) and a large proportion of the total variance attributable to differences between participants (adj. ICC = 0.60).

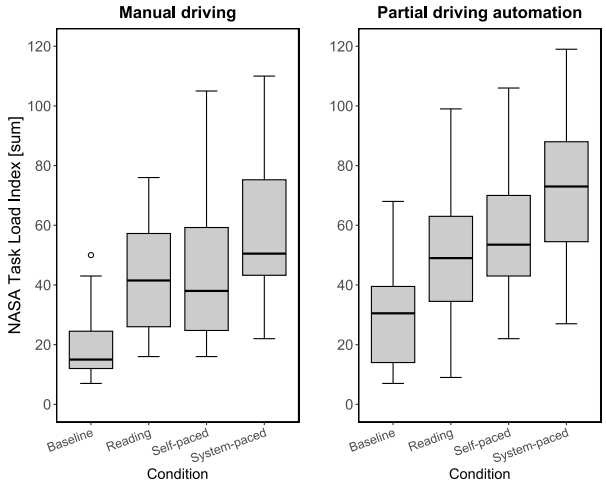


Fig. 3. *NASA-TLX scores*

3.2 Influences on brake response time

Evaluating a surrogate for driving performance (Fig. 4), brake response time in L0 was highest in the system-paced condition ($M = 1.93$, $SD = 0.57$, $n = 25$), while the reading task led to the longest brake response times in L2 ($M = 1.67$, $SD = 0.65$, $n = 31$).

Controlling for driving mode and secondary tasks, demand was positively associated with brake response time, $b = 0.01$, $t(123.96) = 2.74$, $p = .007$, supporting H3. Low variance between participants ($\sigma^2 = 0.05$) accompanied a moderate proportion of total variance attributable to participant-level differences (adj. ICC = 0.20). However, brake response times in L2 were significantly shorter than in L0, $b = –0.40$, $t(51.29) = –4.23$, $p < .001$. Additionally, the reading task significantly increased brake response times, $b = 0.36$, $t(164.98) = 3.58$, $p < .001$ (see Appendix A, Table 2, for details).

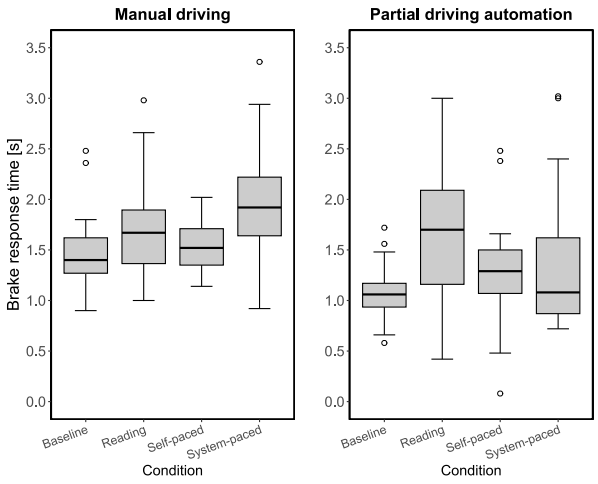


Fig. 4. *Brake response time*

4. Discussion

Results indicate that demand on the driver was influenced by the primary driving task and secondary tasks, with individual differences considered. Higher demand was associated with increased brake response time, controlling for driving mode and secondary tasks. Demand was higher during L2 driving, likely due to the greater effort required to supervise the system. Highest demand was reported for system-paced secondary tasks, corroborating the importance of facilitating driver-paced IVIS interaction. Although demand increased with L2 and secondary tasks, brake response time was not adversely affected. This confirms earlier research that higher demands do not always result in impaired driving performance, as long as driver capacity is not exceeded (Angell et al., 2006; Hurts et al., 2011). However, findings highlight that L2 systems have both positive (Eddine et al., 2024) and negative effects.

Further research is planned to investigate the impact of secondary tasks that cannot be tested with current methods on gaze behaviour and crash risk across levels of driving automation and different metrics for demand and driving performance.

5. Conclusions

The current research provides part of the foundation for developing a test method able to evaluate 'sufficient' attention under distraction (Kircher & Ahlstrom, 2017). It highlights the need to go beyond primarily gaze-based assessments (e.g., NHTSA, 2013) to more completely understand when driver capacities are exceeded, with impaired driving performance and increased crash risk as a result.

6. Acknowledgments

This research was funded by the BMW Group.

References

Angell, L. S., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., & Kiger, S. (2006). *Driver workload metrics task 2 final report* (Final Research Report DOT HS 810 635). Crash Avoidance Metrics Partnership (CAMP).

Biondi, F. N., & Jajo, N. (2024). On the impact of on-road partially-automated driving on drivers' cognitive workload and attention allocation. *Accident Analysis & Prevention*, *200*, 107537. https://doi.org/10.1016/j.aap.2024.107537

Eby, D. W., & Kostyniuk, L. P. (2003). *Driver distraction and crashes: An assessment of crash databases and review of the literature* (Interim Report UMTRI-2003-12). The University of Michigan, Transportation Research Institute. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/1533/97314.0001.001.pdf;sequence=2

Eddine, R., Mulatti, C., & Biondi, F. (2024). On investigating drivers' attention allocation during partially-automated driving. *Cognitive Research: Principles and Implications*, *9*(21). https://doi.org/10.1186/s41235-024-00549-7

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North Holland Press. https://doi.org/10.1016/S0166-4115(08)62386-9

Hurts, K., Angell, L. S., & Perez, M. A. (2011). The distracted driver: Mechanisms, models, and measurement. *Reviews of Human Factors and Ergonomics*, *7*(1), 3–57. https://doi.org/10.1177/1557234X11410387

Kircher, K., & Ahlstrom, C. (2017). Minimum required attention: A human-centered approach to driver inattention. *Human Factors*, *59*(3), 471–484. https://doi.org/10.1177/0018720816672756

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (Technical Report DOT HS 810 594). National Highway Traffic Safety Administration (NHTSA). https://vtechworks.lib.vt.edu/server/api/core/bitstreams/209da0c9-e260-4748-8896-23f6bd14ed01/content

Kujala, T., & Grahn, H. (2022). Key requirements and a method for measuring in-vehicle user interfaces' distraction potential. *The 8th International Conference on Driver Distraction and Inattention*, 60–63.

Kummerer, K., Hergeth, S., Gold, C., & Krems, J. F. (2025). Behavioural components of maintaining a minimum required attention on the roadway during non-driving related tasks that affect safe driving. In A.-C. Hensch, C. Hübner, J. Bornemeier, J. F. Krems, & S. Brandenburg (Eds.), *Proceedings of the 9th Humanist Conference*. https://doi.org/10.60687/2026-0025

Kummerer, K., Hergeth, S., Goreczna, Y., Gold, C., & Krems, J. F. (2026). Mitigating effects of distraction: Drivers adjust speed and glance behaviour during secondary tasks in a driving

simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *119*, 103603. https://doi.org/10.1016/j.trf.2026.103603

Leeuw, J. de, & Meijer, E. (Eds.). (2008). *Handbook of multilevel analysis*. Springer.

Liang, Z., Wang, Y., Qian, C., Wang, Y., Zhao, C., Du, H., Deng, J., Li, X., & He, Y. (2024). A driving simulator study to examine the impact of visual distraction duration from in-vehicle displays: Driving performance, detection response, and mental workload. *Electronics*, *13*(14), 2718. https://doi.org/10.3390/electronics13142718

Metz, B., Schömig, N., & Krüger, H.-P. (2011). Attention during visual secondary tasks in driving: Adaptation to the demands of the driving task. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(5), 369–380. https://doi.org/10.1016/j.trf.2011.04.004

Najm, W. G., & Smith, D. L. (2007, June 18). Definition of a pre-crash scenario typology for vehicle safety research. *Proceedings of the 20th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. https://static.nhtsa.gov/esv/pdf/ESV/Proceedings/20/07-0412-O.pdf

National Highway Traffic Safety Administration (NHTSA). (2013). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices* (NHTSA-2010-0053). National Highway Traffic Safety Administration (NHTSA). https://www.federalregister.gov/documents/2013/04/26/2013-09883/visual-manual-nhtsa-driver-distraction-guidelines-for-in-vehicle-electronic-devices

Regan, M. A., Hallett, C., & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention*, *43*(5), 1771–1781. https://doi.org/10.1016/j.aap.2011.04.008

Society of Automotive Engineers (SAE). (2021). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems* (SAE International Recommended Practice J3016_202104; Version Revised April 2021, Issued January 2014). SAE International. https://doi.org/10.4271/J3016_202104

Strobl, M. H. (2003). SPIDER - The innovative software framework of the BMW driving simulation. *VDI Berichte*, 303–320.

Sundfør, H. B., Sagberg, F., & Høye, A. (2019). Inattention and distraction in fatal road crashes – Results from in-depth crash investigations in Norway. *Accident Analysis & Prevention*, *125*, 152–157. https://doi.org/10.1016/j.aap.2019.02.004

Appendix A

Table 1 *LMM presenting influences on self-reported demand*

Effect	NASA-TLX						
	b	*SE*	*95% CI*		*t*	*p*	*β*
			LL	*UL*			
Fixed effects							
Intercept	19.43	3.81	11.81	26.95	5.10	<.001	-
Partial driving automation (L2) [a]	12.22	4.75	3.18	21.59	2.57	.01	0.49
Reading [b]	19.75	2.58	14.70	24.76	7.66	<.001	0.78
Self-paced [b]	24.49	2.59	19.50	29.58	9.44	<.001	0.97
System-paced [b]	39.80	2.61	34.64	44.79	15.24	<.001	1.58
	σ^2						
Random effect							
Participant	245.0						
Adj. ICC	0.60						

Note. Observations = 197, groups = 51. CI = confidence interval; LL = lower limit; UL = upper limit; adj. ICC = adjusted intraclass correlation.
[a] Reference manual driving (L0). [b] Reference baseline.

Table 2 *LMM presenting influences on brake response time*

Effect	Brake response time						
	b	*SE*	*95% CI*		*t*	*p*	*β*
			LL	*UL*			
Fixed effects							
Intercept	1.34	0.10	1.16	1.53	13.79	<.001	-
NASA-TLX	0.01	0.00	0.00	0.01	2.74	.007	0.25
Partial driving automation (L2) [a]	–0.40	0.10	–0.59	–0.22	–4.23	<.001	–0.71
Reading [b]	0.36	0.10	0.16	0.56	3.58	<.001	0.64
Self-paced [b]	-	-	-	-	-	ns	-
System-paced [b]	-	-	-	-	-	ns	-
	σ^2						
Random effect							
Participant	0.05						
Adj. ICC	0.20						

Note. Observations = 194, groups = 51. CI = confidence interval; LL = lower limit; UL = upper limit; ns = not significant; adj. ICC = adjusted intraclass correlation.
[a] Reference manual driving (L0). [b] Reference baseline.

Validation of an evidence accumulation model for driver readiness in assisted driving using driver ocular metrics

Rafael C. Goçalves [1*], Jorge Pardo Gaytan [3], Jonny Kuo [2], Gustav Markkula [1], Ronald Schroeter [3], Mike G. Lenné [2], Natasha Merat [1]

[1] University of Leeds | Institute for Transport Studies, Leeds, United Kingdom
[2] Seeing Machines Ltd., Melbourne, Australia
[3] Queensland University of Technology, Brisbane, Australia
[*]Trarg@leeds.ac.uk

Abstract: This study tested the validity of driver readiness as an evidence accumulation bias reflecting available cognitive and motor resources, and examined its association with Driver Monitoring System (DMS)-based ocular metrics. Thirty three participants experienced a safety-critical takeover during assisted driving in a simulated motorway driving scenario. The effect of cognitive load, visual occlusion and different levels of fatigue on the quality and timing of this takeover response was assesed during this mixed-participant group study. A single-alternative evidence accumulation model parameterized readiness as a bias factor, derived from combinations of DMS gaze metrics and was compared to a standard model without state information using AICc. The best-fitting model relied on off-road gaze time and stationary gaze entropy, significantly outperforming the reference model (ΔAICc = 7.72, p = .021) and reducing prediction residuals by 10.5% (0.053 s). Results support readiness as a theoretically valid construct linked to gaze behavior, although practical safety gains remain modest.

1. Introduction

Partially-automated vehicles (SAE, 2021) require drivers to monitor the vehicle and road environment, being ready to intervene, if necessary. Driver Monitoring Systems (DMS) address safety issues associated with the removal of the driver as an active controller of the vehicle (Merat et al., 2019), by warning them if drivers this monitoring falls below a certain threshold. The adequancy of drivers' state to manage the safety demands of a particular takeover situation is referred to as drtiver "readiness" or "availability" (Marberger et al., 2017).

While the exact definition of readiness remains a topic of debate in the literature (see Gonçalves et al., 2025a), previous research supports the idea that readiness can be conceptualised as the accumulation of cognitive and motoric resources available to drivers, for resuming control of the driving task (Kim et al, 2022; ISO 2025). This variable fluctuates over time, and is associated with the likelihood of resuming control in a given situation, within a certain time period (Mioch et al., 2017).

The plausibility of this idea is supported by an recently developed model proof of concep (see Gonçalves et al., 2024). Here, we successfuly modelled the variability in drivers' takeover reaction times during a driving simulator study, based on DMS-based ocular metrics, using an evidence accumulation model (EAM, Ratcliff et al., 2016). Drivers' resources were manipulated by the usage of a heavy fog, and a quiz task. This work supported the association between drivers' gaze patterns and the theoretical construct of readiness, showing that a "more ready" driver required the accumulation of less resources, before reaching a timely and appropriate takeover decision (Gonçalves et al., 2024).

To date, most of the empirical work supporting this idea (e.g., Kim et al., 2022; Gonçalves et al., 2024) has not systematically varied the constructs that affect driver state and therefore their readiness. This includes changes in driver state due to fatigue (Guo et al., 2021); cognitive load (Goodridge et al., 2026); and visual distraction (Zeeb et al., 2015). As each of these driver states is generally associated with particular ocular metrics, it is unclear whether this theoretical framework can represent a wide range of driver states, and how they can be detected by different DMS-based ocular metrics.

Based on this research gap, the objective of the present study was to test the validity of the concept of readiness as an accumulation of resources towards a takeover response, for drivers under different driver states, and its association with different DMS-based ocular metrics. Based on this research gap, we posed the following research questions:

- Can the level of driver readiness, as an accumulation of resources, be associated with their capability to safely resume control of the driving task?
- What metrics of a DMS are correlated with drivers' readiness?

2. Method

Drivers' response to a safety-critical scenario was modelled using a single-alternative evidence accumulation model (Rattcliff & Smith, 2004), based on the combination of several ocular metrics extracted from a Seeing Machines PC-DMS eye tracking system, at a capture rate of ~50 hz.

2.1 The model

The model was based on the proof of concept developed by Gonçalves et al. (2024). See the formulae used below:

$$Z = i_a X_a + i_b X_b \quad (1)$$

$$E_t = Z + \sum_{s=0}^{t} \varepsilon_s \quad (2)$$

$$\varepsilon_s \sim N(\mu, \sigma^2), i.i.d. \quad (3)$$

In this model, readiness is represented in (1) as the bias (Z) towards the decision during the evidence accumulation process. This bias is derived from two ocular metrics (X_a, X_b), weighted by their respective constants (i_a, i_b).

In (2), E_t denotes the accumulated evidence at a given time t, and the decision threshold was fixed at 1. In (2) and (3), ε_s represents the noise in the accumulation process, generated from a Gaussian distribution with mean μ and variance σ^2.

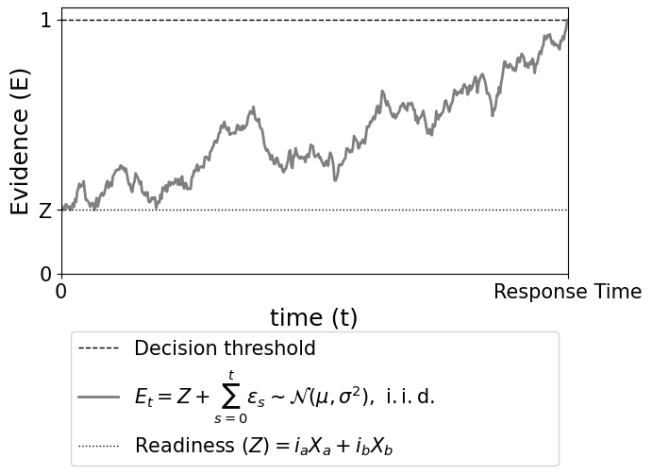


Fig 1 *Graphical representation of the model's concept.*

2.2 Dataset

The model was fit to a driving simulator experiment dataset (see details in Gonçalves et al., 2025b). The experimental scenario was composed of a 3-lane motorway, with SAE L2 assisted driving engaged (SAE, 2021). Participants were required to resume control from the automation due to a system malfunction, leading to a forward-collision critical scenario at a time-to-collision of 3 s.

Drivers' state was manipulated by a cognitively-loading 2-back task (Mehler et al., 2011); a visually-distracting intermittent occlusion manipulation (Kujala et al., 2023), or a combination of both. Drivers' state was also examined as a between-subject factor, since 11 out of the 33 particpants were night-shift workers, who performed the experiment right after their shift.

2.3 DMS ocular metrics selection

The DMS ocular metrics used for readiness calculation were those previously associated with one of our driver state manipulations. See **Table 1**:

Table 1 List of DMS ocular considered.

DMS Ocular Metric	Reference
Gaze time on road (s)	Euro NCAP (2023)
Gaze time driver mirror (s)	Euro NCAP (2023)
Gaze time off road (s)	Euro NCAP (2023)
Gaze time instrument cluster (s)	Euro NCAP (2023)
Gaze transition rate towards on road (percentage)	Schenebelen et al. (2019)
Gaze transition rate towards driver mirror (percentage)	Schenebelen et al. (2019)
Gaze transition rate towards off road (percentage)	Schenebelen et al. (2019)
Gaze transition rate towards instrument cluster (percentage)	Schenebelen et al. (2019)
Vertical gaze dispersion (degrees)	Engström et al. (2017)
Horizontal gaze dispersion (degrees)	Engström et al. (2017)
PERCLOS (percentage)	Kuo et al. (2019)
Blink amplitude/velocity ratio (ms)	Johns (2003)
Gaze stationary entropy (percentage)	Shifferaw et al. (2019)

2.4 Fitting process

The model parametrization process was conducted through a random search (Bergstra & Bengio, 2012). A total of 10,000 search iterations were performed for every possible combination of DMS metrics. The results only report the combination that yielded the best model fit.

2.5 Validation process

The validity of the construct and adequacy of the DMS ocular metrics were assessed by comparing the model's fit to a reference model in the form of the standard evidence accumulation model (Ratcliff et al., 2016), using Akaike Information Criteraion tests (AICc, Burnham & Anderson, 2004). The standard evidence accumulation model was parametrised using the same methods as the study's model.

The reference model uses a constant bias Z, i.e., it has no information on drivers' state, instead modelling inter-trial variability as a random factor. Therefore, if incorporating specific DMS ocular metrics improves model fit, it suggests that these metrics capture meaningful variability in drivers' takeover performance, indicating a relationship with the underlying construct of readiness. Consequently, significant improvements in model fit would provide evidence supporting the proposed concept of readiness as an accumulation of resources influencing takeover responses.

3. Results

The best performing model calculated its bias Z as a weighted sum of off road gaze time and stationary gaze entropy. The AICc test results showed that this model significantly outperformed the reference model (ΔAICc = 7.717, p = 0.021), representing an average 10.49% reduction on the prediction residuals (0.053 s). See **Fig 2** and **Fig 3** for the performance comparisons and distribution of model residuals.

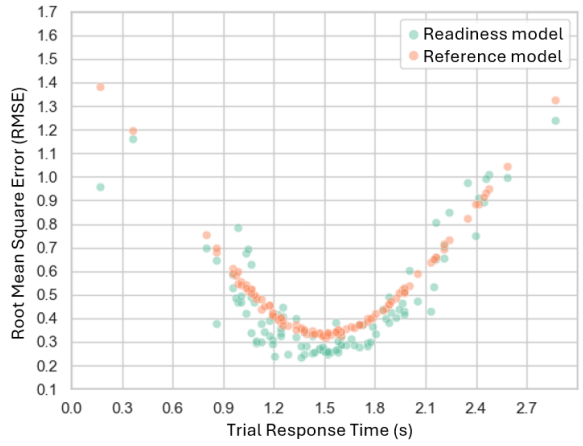


Fig 2 *Model performance comparison.*

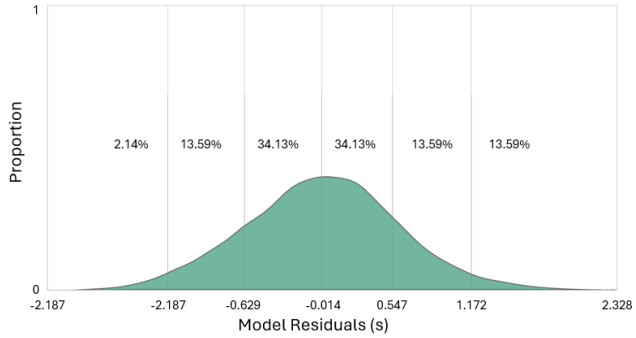


Fig 3 *Model residual distribution.*

4. Discussion

The best-performing model variant used off road glance-based metrics, reiterating the importance of the focus on visual distraction from current DMS recommendations (Euro NCAP, 2023; UNECE, 2024). The AICc results supports the theoretical validity of readiness as an accumulation of resources, that can be used to explain the range of likely responses of participants. The residual improvements suggests that readiness estimators seem to explain a small, but significant part of the phenomenon. Although the improvement was statistically significant, the model's residuals remain too large for reliable real-world application, ranging from -0.62 to 0.55 s. Moreover, given the magnitude of the observed effects (10.49%), the gain in predictive accuracy is unlikely to translate into a meaningful reduction in crash outcomes if applied to actual DMS solutions.

5. Conclusions

While gaze metrics metrics are promising readiness indicators, more complex relationships need to be explored with further research. This suggests that the proposed readiness concept is theoretically related valid, but may not be the only factor affecting drivers' takeover performance.

6. Acknowledgments

The experiment reported in this work was funded by the University of Leeds and Seeing Machines Ltd. Collaboration Fellowship.

References

Bergstra, J., Ca, J. B., & Ca, Y. B. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. In Journal of Machine Learning Research (Vol. 13). https://doi.org/10.5555/2188385.2188395

Burnham, K. P., & Anderson, D. R. (2016). Multimodel Inference: Understanding AIC and BIC in Model Selection. Http://Dx.Doi.Org/10.1177/0049124104268644, 33(2), 261– 304. https://doi.org/10.1177/0049124104268644

Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. Human Factors, 59(5), 734–764.

Euro NCAP. (2023). Assisted driving highway assist systems (Technical report). Euro NCAP.

Gonçalves, R. C., Goodridge, C. M., Kuo, J., Lenné, M. G., & Merat, N. (2024). Using driver monitoring to estimate readiness in automation: A conceptual model based on simulator experimental data. Cognition, Technology & Work, 26(4), 569–584. https://doi.org/10.1007/s10111-024-00777-3

Gonçalves, R. C., Goodridge, C. M., Pardo, J., Reher, A., Kuo, J., Bruneau, A., & Merat, N. (2025, September). What makes a ready driver? A deep dive into the measurement and validation of readiness estimation for driver monitoring systems of partially automated (SAE Level 2) vehicles [Workshop]. In Adjunct Proceedings of the 17th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM.

Gonçalves, R. C., Pardo, J., Elhenawy, M. M. Z., Kuo, J., Azarmi, M., Rezaei, M., ... & Merat, N. (2025, September). It Matters Who Is Behind The Wheel: Driver Monitoring Feature Analysis Using Explainable AI. In Adjunct Proceedings of the 17th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 245-249).

Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., Bruneau, A., Lee, Y. M., & Merat, N. (2026). The impact of N-back-induced mental workload and time budget on takeover performance. Accident Analysis & Prevention, 225, 108327. https://doi.org/10.1016/j.aap.2025.108327

Guo, H., Zhang, Y., Cai, S., & Chen, X. (2021). Effects of level 3 automated vehicle drivers' fatigue on their take-over behaviour: A literature review. Journal of advanced transportation, 2021(1), 8632685.

International Organization for Standardization. (2025). ISO/TS 5283-1:2025 — Road vehicles — Driver readiness and intervention management — Part 1: Partial automation (Level 2). ISO.

How L3 Driving Automation Will Actually Be Used: A Phase-Based Taxonomy of Usage Patterns and Design Considerations

Joonbum Lee[1*], John D. Lee[1], Joseph F. Szczerba[2], Akilesh Rajavenkatanarayanan[2], and Roy Mathieu[2]

[1] Department of Industrial and Systems Engineering, University of Wisconsin–Madison, USA

[2] General Motors Global Research & Development, USA

[*] joonbum.lee@wisc.edu

Abstract

SAE Level 3 (L3) driving automation permits complete driver disengagement, allowing drivers to engage in non-driving related tasks while demanding rapid takeover capability when the system reaches its operational limits. As L3 features are introduced alongside existing L1 and L2 platforms, drivers will frequently transition between modes, experiencing repeated cycles of disengagement and re-engagement that create challenges for distraction and inattention management. Yet no systematic framework characterizes how these usage patterns shape the human factors demands on drivers and interfaces. We present a phase-based factorial taxonomy organizing L3 usage into Entry, During, and Exit phases across five dimensions, yielding 48 distinct patterns. Duration emerges as the primary determinant of human factors challenges, with eight extended-duration, system-initiated patterns requiring the greatest HMI design attention. The taxonomy provides a structured basis for designing interfaces matched to pattern-specific demands, from mode confusion during brief transitions to the sustained inattention inherent in extended automation use.

1. Introduction

SAE Level 3 (L3) conditional driving automation represents a fundamental shift in driver responsibility. Unlike L2 systems requiring continuous monitoring, L3 permits complete driver disengagement—drivers may direct attention entirely to non-driving related tasks (NDRTs) such as reading, watching video, or using mobile devices—while demanding rapid takeover capability when the system reaches its operational limits (SAE International, 2021). This tension between disengagement and readiness creates human factors challenges not addressed by interfaces designed for lower automation levels.

Research has identified several core challenges. Drivers become progressively out-of-the-loop (OOTL) during automation, losing situation awareness, experiencing vigilance decrement, and accumulating passive fatigue (Endsley, 2017). Takeover performance degrades with longer automation exposure (Eriksson & Stanton, 2017), and NDRT engagement—a form of driver distraction sanctioned by the system—further impairs takeover readiness (Radlmayr et al., 2014). Mode confusion between L2 and L3 compounds the problem: vehicle behaviour appears identical while monitoring responsibilities differ fundamentally (Eom & Lee, 2022).

Despite extensive research, studies have largely treated L3 takeover events in isolation without considering the broader usage context. In practice, L3 will function as an incremental feature alongside L1, L2, and manual driving, with frequent mode transitions as operational conditions change. We address this gap with a phase-based taxonomy that systematically characterizes L3 usage episodes and identifies patterns that create the most demanding conditions for driver attention, including those where sustained inattention makes reorientation most difficult.

2. Method

2.1 Phase-based framework

We conceptualize each L3 usage episode as a temporal sequence of three phases, each associated with distinct attentional demands:

1. **Entry:** Transition into L3. The driver disengages from the driving task and redirects attention to NDRTs.
2. **During:** Period of L3 engagement. The driver is disengaged from driving, with inattention to the road environment accumulating over time.
3. **Exit:** Transition out of L3. The driver must reorient attention from NDRTs back to the driving task.

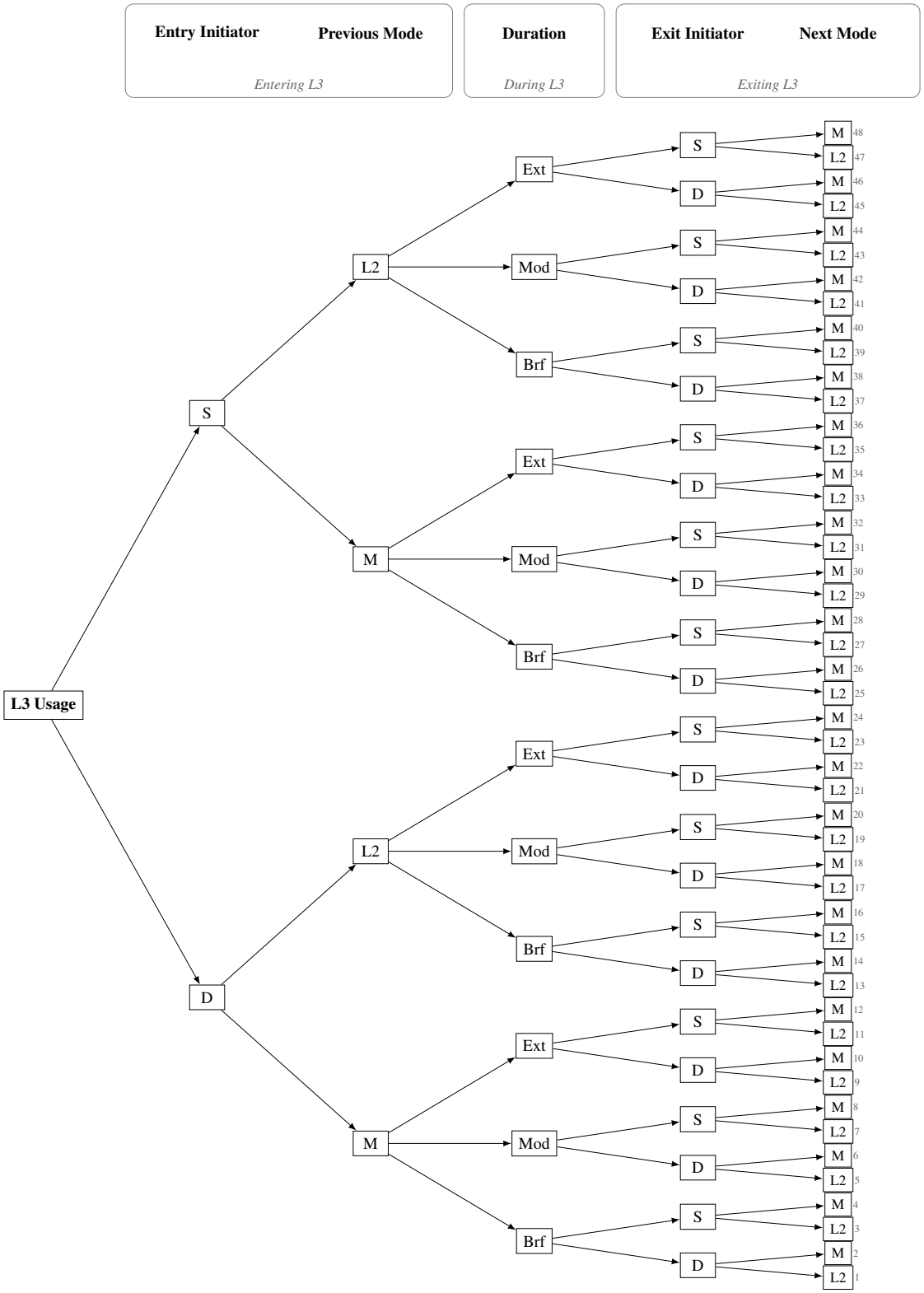


Figure 1: *Complete taxonomy tree: 48 L3 usage patterns ($2\times2\times3\times2\times2$). Phases are labelled above: Entering L3, During L3, Exiting L3. Reading left to right: Entry Initiator (D=Driver, S=System), Previous Mode (M=Manual, L2), Duration (Brf=Brief, Mod=Moderate, Ext=Extended), Exit Initiator (D, S), Next Mode (L2, M). Numbers 1–48 identify each terminal pattern.*

2.2 Factorial dimensions

We identified five dimensions through analysis of L3 implementations and human factors literature. Table 1 summarizes each dimension.

Table 1: *Factorial dimensions of L3 usage patterns.*

Dimension	Levels	HMI Relevance
Entry Initiator	Driver, System	System-initiated entry may go unnoticed; mode change without driver action raises mode confusion concerns (Lu & de Winter, 2015)
Previous Mode	Manual, L2	L2→L3 lacks behavioural cues (vehicle behaviour appears identical) despite fundamentally different responsibilities. L1 is subsumed under Manual for parsimony
Duration	Brief, Moderate, Extended	Determines driver state during automation; nonlinear relationship with takeover performance (Portron et al., 2024)
Exit Initiator	Driver, System	System-initiated exits represent mandatory transitions at operational boundaries
Next Mode	L2, Manual	L3→Manual requires full control resumption; L3→L2 demands monitoring while vehicle appears unchanged

The full crossing yields $2 \times 2 \times 3 \times 2 \times 2 = 48$ distinct usage patterns (Figure 1). Each pattern is encoded as `[EntryInit]-[PrevMode]|[Dur]|[ExitInit]-[NextMode]`; for example, `D-M|Ext|S-L2` denotes driver-initiated entry from manual, extended duration, system-initiated exit to L2.

Duration levels are informed by Portron et al. (2024), who found a nonlinear relationship between L3 duration and takeover performance: 15 min produced the best performance (0% crash rate), 5 min showed moderate crash rates (15%), and 45 min the highest (23–31%), despite drowsiness increasing linearly. We define three levels (Table 2):

Table 2: *Duration levels and associated driver states.*

Duration	Driver State	Takeover Characteristics
Brief (<5 min)	Adjusting to automation; calibrating trust, transitioning into NDRTs	Moderate crash rate (15% at 5 min); driver not yet fully disengaged
Moderate (5–15 min)	Adapted but limited OOTL effects; stable NDRT engagement	Best takeover performance (0% crash rate at 15 min); mode confusion risk at L3→L2 exits
Extended (>15 min)	Progressive OOTL: SA degradation, vigilance decrement, passive fatigue	Highest crash rates (23–31% at 45 min); deep NDRT absorption

Note. Crash rates from Portron et al. (2024).

2.3 Pattern icon

To support rapid pattern comparison, we developed a Lego-block icon (Figure 2) encoding each pattern's dimensions as interlocking blocks in a bird's-eye layout. Block colour indicates automation level (teal=L3, green=L2, yellow=Manual). Transition block height encodes the mode gap (taller = larger gap), and colour encodes the initiator (white=system, gray=driver).

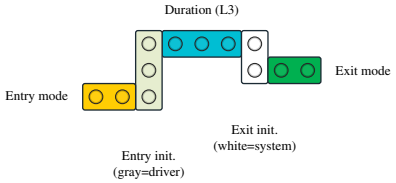


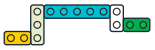
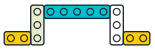
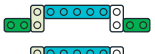
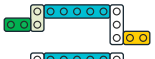
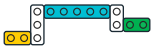
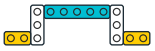
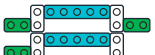
Figure 2: *Lego-block pattern icon (example: D-M—Mod—S-L2). Colour encodes automation level: yellow=Manual, green=L2, teal=L3. Duration width scales with time (1/3/5 studs). Transition height encodes mode gap; colour encodes initiator (white=system, gray=driver).*

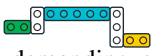
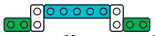
3. Results

3.1 Highest-priority patterns

Eight extended-duration patterns with system-initiated exits present the greatest demands for attention recapture (Table 3). These patterns share a common structure: extended inattention followed by an externally triggered demand to reorient attention—either to active monitoring (L3→L2) or full vehicle control (L3→Manual).

Table 3: *Eight highest-priority patterns for HMI design.*

#	Pattern	Icon	Primary Concern
11	D-M—Ext—S-L2		Extended OOTL + mode confusion
12	D-M—Ext—S-M		Extended OOTL + urgent takeover
23	D-L2—Ext—S-L2		L2 origin + max mode confusion
24	D-L2—Ext—S-M		L2 origin + urgent takeover
35	S-M—Ext—S-L2		Passive entry + mode confusion
36	S-M—Ext—S-M		Passive entry + urgent takeover
47	S-L2—Ext—S-L2		Full L2/L3 cycling
48	S-L2—Ext—S-M		Passive, extended, urgent takeover

Pattern 48  (`S-L2|Ext|S-M`) combines the most demanding conditions: automatic L2-to-L3 upgrade, extended disengagement, then system-initiated exit requiring immediate manual takeover. Pattern 47  (`S-L2|Ext|S-L2`) presents another demanding mode confusion challenge: both transitions are system-initiated, vehicle behaviour is identical throughout, and the driver must shift from complete disengagement to active monitoring with no distinguishing cue.

3.2 HMI design implications

The taxonomy analysis yields three categories of design principles for L3 communication systems.

Continuous mode awareness: L2 and L3 must be continuously and unambiguously distinguishable without active display checking. Because vehicle behaviour

appears identical, ambient channels (e.g., distinct interior lighting colours) can maintain peripheral mode awareness, with audio and haptic channels reinforcing transitions (Mirnig et al., 2023). This is particularly important given that drivers engaged in NDRTs may be inattentive to mode state.

Duration-adaptive communication: Communication strategy should adapt to engagement duration. Brief patterns require clear mode confirmation at entry and rapid re-orientation support at exit; moderate patterns need advance preview for system-initiated exits; extended patterns require periodic minimal interaction to prevent complete disengagement and progressive escalation toward transitions (Politis et al., 2015).

Transition management: All system-initiated transitions should provide maximum preview time with graduated escalation. Transitions to L2 require clear mode differentiation and monitoring reminders; urgent manual takeovers require simultaneous activation of all modalities at maximum intensity. Haptic cues (e.g., seat vibration) are critical for capturing attention from NDRTs, particularly after extended durations where drivers may be deeply absorbed and visual and auditory cues alone may be insufficient (Gold et al., 2016).

4. Discussion

The taxonomy situates each L3 transition within its surrounding usage context—how the driver entered L3, how long they have been disengaged, and what the exit demands. This contextual perspective reveals that the same exit event (e.g., system-initiated transition to manual) creates fundamentally different challenges depending on whether it follows 3 minutes or 45 minutes of disengagement, during which drivers may become progressively inattentive to the driving environment.

A practical implication is that L3 will function as an incremental feature added to existing L1 and L2 platforms, implying frequent mode transitions as operational boundaries shift. These transitions create repeated cycles of disengagement and re-engagement—a pattern that conventional interfaces, designed for sustained monitoring, do not address. The taxonomy provides a systematic basis for designing context-adaptive interfaces that modulate communication intensity based on the current usage pattern, delivering stronger interventions for the most demanding patterns while avoiding unnecessary interruption during lower-demand episodes.

The framework does not capture all moderating factors—NDRT type, individual differences, and environmental conditions further shape driver state dynamics. However, the factorial structure accommodates future refinement through nested factors without requiring changes to the core dimensions. Future work will use this taxonomy to design and evaluate context-adaptive communication solutions matched to pattern-specific demands.

5. Conclusions

This phase-based taxonomy provides a systematic framework for characterising how L3 driving automation will be used in contexts where drivers transition between manual, L1, L2, and L3 modes. The $2 \times 2 \times 3 \times 2 \times 2$ factorial structure yields 48 usage patterns, with duration as the primary determinant of human factors challenges and eight extended-duration, system-initiated patterns presenting the greatest demands for HMI design. The framework offers a structured basis for designing interfaces that support drivers across the full range of L3 usage contexts, addressing challenges from mode confusion to the distraction and inattention inherent in extended automation use.

Acknowledgements

This research is supported by General Motors.

References

Endsley, M. R. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making*, 11(3), 225–238.

Eom, H., & Lee, J. D. (2022). Mode confusion of human–machine interfaces for automated vehicles. *Journal of Computational Design and Engineering*, 9(5), 1995–2009.

Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*, 59(4), 689–705.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2016). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 1938–1942.

Lu, Z., & de Winter, J. C. F. (2015). A review and framework of control authority transitions in automated driving. *Procedia Manufacturing*, 3, 2510–2517.

Mirnig, A. G., Gärtner, M., Wallner, V., et al. (2023). Enlightening mode awareness: Assisting drivers of conditionally automated vehicles with peripheral ambient light displays. *Personal and Ubiquitous Computing*, 27, 2307–2320.

Politis, I., Brewster, S., & Pollick, F. (2015). Language-based multimodal displays for the handover of control in autonomous cars. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 3–10.

Portron, A., Perrotte, G., Ollier, G., Bougard, C., Bourdin, C., & Vercher, J.-L. (2024). Getting back in the loop: Does autonomous driving duration affect driver's takeover performance? *Heliyon*, 10, e24112.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Fac-*

tors and Ergonomics Society Annual Meeting, 58(1), 2063–2067.

SAE International. (2021). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles* (SAE Standard J3016_202104).

Kim J, Kim W, Kim HS, Lee SJ, Kwon OC, Yoon D (2022) A novel study on subjective driver readiness in terms of non-driving related tasks and take-over performance. ICT Express 8(1):91–96. https://doi.org/10.1007/978-3-319-60441-1_58/ COVER

Kujala, T., Kircher, K., & Ahlström, C. (2023). A Review of Occlusion as a Tool to Assess Attentional Demand in Driving. Human Factors, 65(5), 792–808. https://doi.org/10.1177/00187208211010953/ASSET/IMAGES/LARGE/10.1177_00187208211010953-FIG2.JPEG

Kuo, J., Lenné, M. G., Mulhall, M., Sletten, T., Anderson, C., Howard, M., ... & Collins, A. (2019). Continuous monitoring of visual distraction and drowsiness in shift-workers during naturalistic driving. Safety Science, 119, 112-116.

Marberger C, Mielenz H, Naujoks F, Radlmayr J, Bengler K, Wandtner B (2018) Understanding and applying the concept of "Driver availability" in automated driving. Adv Intell Syst Comput 597:595–605. https://doi.org/10.1007/978-3-319-60441-1_58/ COVER

Mehler, B., Reimer, B., & Dusek, J. A. (2011). MIT AgeLab Delayed Digit Recall Task (n-back)(Paper 2011-3B)(2011-06-28).

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2019). The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications. Cognition, Technology and Work, 21(1), 87–98. https://doi.org/10.1007/s10111-018-0525-8

Mioch T, Kroon L, Neerincx M (2017) Driver readiness model for regulating the transfer from automation to human control figure 1. Truck platooning experiment with a CTS agent. Proceedings of the 22nd International Conference on Intelligent User Interfaces. https://doi.org/10.1145/3025171

Ratcliff R, Smith PL (2004) A comparison of sequential sampling models for two-choice reaction time. Psychological review, 111(2):333.

Ratcliff R, Smith PL, Brown SD, McKoon G (2016) Diffusion decision model: current issues and history. Trends Cogn Sci 20(4):260– 281. https://doi.org/10.1016/J.TICS.2016.01.007

SAE. (2021). Summary Of SAE International'S Levels Of Driving Automation For On-Road Vehicles. https://www.sae.org/misc/pdfs/automated_driving.pdf

Schnebelen, D., Charron, C., & Mars, F. (2019). Predicting self-assessment of the out-of-the-loop phenomenon from visual strategies during highly automated driving. In Proceedings of the Human Factors and Ergonomics Society Europe Chapter Annual Meeting 2019 (pp. 29–39).

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. Neuroscience & Biobehavioral Reviews, 96, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007

United Nations Economic Commission for Europe. (2024, October 4). New UN regulation paves way for deployment of driving assistance systems worldwide [Press release]. UNECE. https://unece.org/media/press/395206

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. Accident Analysis & Prevention, 78, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

From Old Bottles to New: Reframing Driver Distraction in an Era of Vehicle Automation

Joonbum Lee[1*] and John D. Lee[1]

[1] Department of Industrial and Systems Engineering, University of Wisconsin–Madison, USA

[*] joonbum.lee@wisc.edu

Abstract

Driver distraction persists despite decades of research, education, and enforcement. Traditional frameworks define distraction as attention diverted from driving by secondary activities; a framing that served manual driving well. Vehicle automation, however, transforms what driving means and what distraction means. We argue distraction must be reconsidered as a joint property of driver, situation, and vehicle technology rather than an internal driver state. This reconceptualization has three implications: (a) supervision requires nested situation awareness of both environment and automation; (b) boundaries between primary and secondary tasks blur, invalidating traditional metrics such as the 2-second glance threshold; (c) inattention inadequately captures disengagement, which reflects task perseveration and attention traps. This framework offers directions for research, design, and policy.

1. Introduction

As the biblical parable warns, "No one puts new wine into old wineskins. For the old skins would burst from the pressure, spilling the wine and ruining the skins" (Matthew 9:17). This paper argues that applying traditional distraction frameworks to automated driving is precisely this mistake—putting new wine in old bottles that cannot contain the new reality.

For decades, driver distraction has been understood through a clear definition: "the diversion of attention away from activities critical for safe driving toward a competing activity" (Lee et al., 2008). This framework served manual driving research well, where the primary task—controlling the vehicle—was unambiguous.

Vehicle automation transforms this landscape. With SAE Level 2 partial automation, drivers shift from active control to supervisory monitoring. With Level 3 conditional automation, drivers may legally disengage entirely within the operational design domain. A single trip may span multiple levels as conditions change, and distraction takes on different meanings: behavior appropriate during Level 3 becomes dangerous during Level 2 and critical during manual control.

Despite decades of research, education, and enforcement, distraction persists. In 2023, distraction-affected crashes claimed 3,275 lives in the United States, representing 8% of all fatal crashes (NHTSA, 2024). This persistence suggests traditional approaches—treating distraction as a driver problem—may be insufficient.

We argue that distraction must be understood as a joint property of driver, road situation, and vehicle technology rather than an internal driver state. Following Hollnagel & Woods (1983), who argued that cognitive systems engineering requires treating human-machine as a joint cognitive system rather than separate interacting components, we develop this argument by establishing the conceptual framework, examining how automation blurs task boundaries, arguing that inattention inadequately captures disengagement, and reviewing statistical methods for characterizing heterogeneous behavior.

2. Reframing Distraction as a Joint Property

Traditional distraction research treats the driver as the problem, assuming stable demands and driver-controlled attention. This emphasis on attentional resources (capacity limits) neglected attentional dynamics (engaging and disengaging) (Lee, 2014). Vehicle automation demands reconsideration: technology affects both resources (by freeing capacity) and dynamics (by changing engagement patterns).

Automation changes the demand side. Bainbridge (1983) identified the "ironies of automation": by removing easy parts of a task, automation makes difficult parts harder. Woods (1997) showed that "clumsy automation" provides benefits during low-workload periods but imposes costs during high-workload moments. For driving, attentional demands are lowest when automation performs well and highest when it fails—precisely when drivers are least prepared.

Moreover, automation itself becomes a distraction

source, introducing tasks absent in manual driving: monitoring mode status, interpreting system messages, and managing transitions. These supervisory tasks compete for the same resources as traditional secondary activities.

With SAE Level 2, supervision encompasses awareness of both road environment and driving automation (Lee et al., 2026). This nested situation awareness (Figure 1) changes where drivers look. Displays convey mode information and environmental data; detected vehicles appear on instrument panels. Understanding of the environment becomes mediated by understanding of automation's perception.

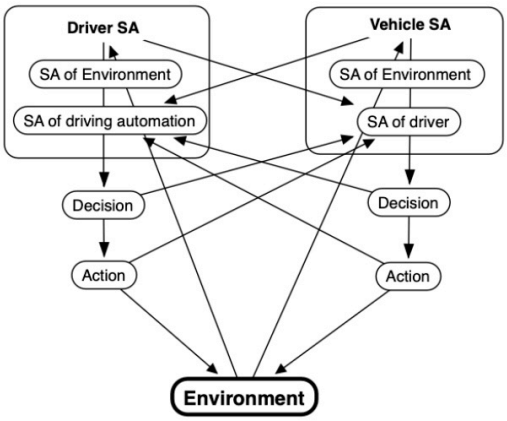


Figure 1: *Nested situation awareness framework for supervising automated driving. The driver needs SA of both the environment and the driving automation; the vehicle needs SA of both the environment and the driver. Modified from Lee et al. (2026).*

Road demands must be considered alongside technology. Senders et al. (1967) characterized demands by information density, which drivers sample intermittently. ADAS features mitigate these demands; a winding road requires less attention with lane centering active. Distraction thus depends on attention relative to demands given automation capability.

Drivers bring trust and mental models, but technology shapes these. Complex systems make accurate mental models difficult. Overtrusting drivers allocate insufficient attention; undertrusting drivers allocate excess. Calibration is difficult when automation varies unpredictably.

3. Blurry Boundaries: Primary and Secondary Tasks

This joint framework requires reexamining traditional boundaries. The distinction between primary and secondary tasks becomes ambiguous; activities once secondary (checking displays) can serve supervision. With SAE Level 3, the primary task may have nothing to do with driving. Distraction as diverted attention loses coherence when the primary task is absent.

The boundary between on-road and off-road glances similarly dissolves. In manual driving, studies applied binary classification with metrics such as percent road center. The AttenD algorithm advanced beyond binary classification by treating mirror/speedometer glances as relevant only under one second and tracking glance history through an attention buffer (Kircher & Ahlström, 2009). Even this classification becomes problematic when displays convey automation status.

With partial automation, drivers monitor status through instrument clusters and screens. Glances to these displays serve situation awareness, yet traditional metrics classify them as off-road. Information once available only through the windshield now appears on displays. However, not all in-vehicle glances serve supervision; troubleshooting infotainment produces distracting glances even at vehicle displays.

NHTSA's Visual-Manual Guidelines (NHTSA, 2013) recommend limiting off-road glances to 2 seconds. Automation changes this: Victor et al. (2018) found that during reliable automation, all participants exceeded 2-second off-path glances, with some reaching 40 seconds. The question shifts from "How long did the driver look away?" to "Did the driver maintain sufficient awareness to intervene?"

4. Beyond Inattention: Capturing Driver Disengagement

Blurring task boundaries suggests traditional constructs are inadequate. Inattention encompasses distraction, fatigue, and other failures, but proves insufficient for characterizing disengagement during automated driving.

Inattention implies a binary state: attending adequately or not. Drivers are often disengaged without secondary tasks; mind wandering is prevalent (Lee, 2014). Automation may exacerbate disengagement by reducing moment-to-moment demands. Automated driving demands nuanced calibration: neither over-monitoring nor under-monitoring.

Disengagement encompasses multiple dimensions: comprehension of automation status, anticipation of behavior, and preparation for intervention. A driver may maintain visual attention yet fail to comprehend that automation is approaching limits.

Disengagement also reflects breakdowns in interruption management (Lee, 2014). Task perseveration occurs when drivers fixate on completing secondary activities. Three factors influence perseveration (Fox & Hoffman, 2002): proximal closure, goal emergence, and goal valence. With Level 2, perseveration conflicts with constant takeover readiness.

Drivers employ distinct glance strategies (Lee et al., 2015). With Level 2, fewer but longer glances prove hazardous as takeover demands arise unpredictably. Individual differences mean aggregate measures obscure different risk profiles.

Three attention traps explain why drivers engage in distracting activities (Lee, 2014): contingency traps

(hazards difficult to perceive), consequence traps (rewards outweigh costs), and conditioning traps (incident-free history creates habits). Automation deepens each trap through extended incident-free operation and sanctioned secondary activities.

A Jevons paradox compounds these dynamics (Jevons, 1865). Just as efficiency gains in coal use led to increased consumption, automation's reduction of driving demands does not preserve attentional capacity—drivers consume freed capacity through secondary tasks. Automation efficiency paradoxically increases exposure to the risks it was meant to reduce.

5. Statistical Approaches for Heterogeneous Behavior

Reconceptualizing distraction has methodological implications. Understanding automated driving behavior requires methods capturing heterogeneity. Traditional approaches assuming homogeneous behavior prove inadequate.

Quantile regression characterizes the entire conditional distribution (DinparastDjadid et al., 2021). Safety-critical information resides in distribution tails; modest mean increases might dramatically increase extreme glances. Two-part models address mixed discrete-continuous outcomes (Lee et al., 2026), decomposing glance behavior into occurrence (discrete) and duration (continuous). Mixture models identify subpopulations with distinct patterns (Lee et al., 2024); aggregate relationships may obscure opposite effects in subgroups.

Distributional assumptions also require attention. Glance measures often violate normality, following right-skewed distributions. Beta regression addresses such distributions, and is particularly well-suited for bounded proportion data such as on-road glance percentage (Lee et al., in press).

These methods, however, cannot compensate for measures that no longer capture the intended construct. Metrics designed for manual driving may not validly assess attention when automation transforms what "attending to driving" means.

6. Discussion

This analysis converges on a central insight: distraction during automated driving cannot be addressed through driver-focused interventions alone.

For research, traditional metrics require reconsideration. Studies must assess whether manual driving measures retain validity during automated driving. New measures capturing supervision quality and intervention readiness merit development.

For design, automation that communicates status clearly may reduce disengagement more effectively than warnings targeting driver behavior.

For policy, threshold-based approaches are complicated when off-road glance meaning depends on automation mode. Performance-based approaches may prove more valid than prescriptive rules.

7. Conclusions

Driver distraction has resisted solution for 50 years partly because it is not solely a driver problem. Understanding distraction now requires recognizing it as a joint property of driver capabilities, situational demands, and vehicle technology.

This reconceptualization reveals three challenges. Supervising automation creates nested situation awareness requirements. Blurring task boundaries undermines classifications such as on-road versus off-road glances. The multidimensional nature of disengagement exceeds what "inattention" can capture.

Statistical methods offer tools for characterizing behavioral heterogeneity, but their value depends on measuring the right constructs. Addressing distraction demands both methodological sophistication and conceptual clarity about what distraction means when vehicles increasingly drive themselves.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

DinparastDjadid, A., Lee, J. D., Domeyer, J., Schwarz, C., Brown, T. L., & Gunaratne, P. (2021). Designing for the extremes: Modeling drivers' response time to take back control from automation using Bayesian quantile regression. *Human Factors*, 63(5), 756–767. https://doi.org/10.1177/0018720819893429

Fox, S., & Hoffman, M. (2002). Escalation behavior as a specific case of goal-directed activity: A persistence paradigm. *Basic and Applied Social Psychology*, 24(4), 273–285. https://doi.org/10.1207/S15324834BASP2404_3

Hollnagel, E., & Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies*, 18(6), 583–600. https://doi.org/10.1016/S0020-7373(83)80034-0

Jevons, W. S. (1865). *The coal question: An inquiry concerning the progress of the nation, and the probable exhaustion of our coal-mines*. Macmillan.

Kircher, K., & Ahlström, C. (2009). Issues related to the driver distraction detection algorithm AttenD. In *Proceedings of the First International Conference on Driver Distraction and Inattention* (pp. 1–15). Gothenburg, Sweden.

Lee, J. D., Young, K. L., & Regan, M. A. (2008). Defining driver distraction. In M. A. Regan, J. D. Lee, & K. L. Young (Eds.), *Driver distraction: Theory, effects, and mitigation* (pp. 31–40). CRC Press.

Lee, J. D. (2014). Dynamics of driver distraction: The process of engaging and disengaging. *Annals of Advances in Automotive Medicine*, 58, 24–32.

Lee, J., Reimer, B., Mehler, B., Angell, L., Seppelt, B. D., & Coughlin, J. F. (2015). Analyses of glance patterns of older and younger drivers during a visual-manual HMI interaction. *Transportation Research Record*, 2518, 29–37. https://doi.org/10.3141/2518-04

Lee, J., Gershon, P., Domeyer, J., Mehler, B., & Reimer, B. (2024). Understanding hands-off patterns while using Tesla Autopilot: Individual differences and their impact on hands-off duration. In *Proceedings of the 8th International Conference on Driver Distraction and Inattention*. Gothenburg, Sweden.

Lee, J., Gershon, P., Domeyer, J., Mehler, B., & Reimer, B. (in press). Hands off the wheel: How individual differences shape driver inattention and disengagement in partial driving automation. *Transportation Research Part F: Traffic Psychology and Behaviour*.

Lee, J., Rheem, H., Lee, J. D., Szczerba, J. F., Rajavenkatanarayanan, A., & Mathieu, R. (2026). With a little help from my car: Sharing automated vehicle's situation awareness reduces driver-initiated automation disengagement without delaying takeover response time. *Human Factors*. https://doi.org/10.1177/00187208261422917

National Highway Traffic Safety Administration. (2013). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices* (Docket No. NHTSA-2010-0053). U.S. Department of Transportation.

National Highway Traffic Safety Administration. (2024). *Distracted driving in 2023* (Research Note DOT HS 813 703). U.S. Department of Transportation.

Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record*, 195, 15–33.

Victor, T. W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., & Ljung Aust, M. (2018). Automation expectation mismatch: Incorrect prediction despite eyes on threat and hands on wheel. *Human Factors*, 60(8), 1095–1116. https://doi.org/10.1177/0018720818788164

Woods, D. D. (1997). Coping with complexity: The psychology of human behaviour in complex systems. In L. Bainbridge & S. A. R. Quintanilla (Eds.), *Developing skills with information technology* (pp. 33–94). Wiley.

Euro NCAP 2029 Driver Monitoring Assessment Framework for Mitigating Driver Inattention Risks

Adriano Palao [1*]

[1] European New Car Assessment Programme (Euro NCAP), Mgr Ladeuzeplein 10, Leuven, Belgium
[*]adriano_palao@euroncap.com

Euro NCAP is developing its 2029 assessment framework for Driver Monitoring, building on the 2026 approach. The new framework aims to move beyond a mainly binary interpretation of visual distraction and towards a more context-aware representation of driver inattention risk. A key objective is to encourage a more meaningful use of Driver State Link, allowing vehicle warning and intervention strategies to better reflect the condition of the driver, while ensuring robust behaviour whenever driver state cannot be determined reliably. To support this evolution, Euro NCAP is exploring a three-pillar assessment methodology based on Spot Checks, On-road Evaluation and Synthetic Data. Together, these methods aim to improve repeatability, real-world relevance and coverage of Driver Monitoring performance. The overall goal is to promote vehicle systems that support the driver effectively when support is needed, while avoiding unnecessary or poorly timed interventions that may undermine driver acceptance.

1. Introduction

Since the introduction of Driver State Monitoring assessment in 2023 as part of the 5-star rating, Euro NCAP has evaluated to date hundreds of vehicles equipped with systems able to detect and respond to a variety of driver states associated with risk, in particular distraction, drowsiness, (micro)sleep and unresponsiveness. In 2026, a new rating system was introduced along with updated requirements across passive and active safety areas. The 2026 assessment for Driver Monitoring evolved to raise the bar in terms of detection accuracy while preserving driver acceptance, and to maximize the life-saving potential of these systems, particularly by linking the sensitivity of ADAS with the state of the driver.

This paper covers the work currently being developed by Euro NCAP towards defining the 2029 assessment of Driver Monitoring. Built upon the 2026 foundation, the framework primarily focuses on achieving the correct balance between effective driver support and driver acceptance, while ensuring a robust and inclusive driver state detection performance with a three-pillar assessment methodology. It is important to note that the concepts described here are work in progress, hence subject to change before the final implementation effective in 2029.

2. Criteria

2.1 Driver Monitoring within the new Euro NCAP Rating Scheme

Within the 2026 Euro NCAP rating scheme, Driver Monitoring sits under Driver Engagement in the Safe Driving stage. The Driver Monitoring assessment is structured around two groups of driver states: transient states such as distraction, and non-transient states such as impairment. This overall structure will be retained for 2029. However, the underlying framework is expected to evolve in order to better reflect how driver attention, driver understanding and external driving context interact in real traffic.

2.2 Driver State Link

The triggering conditions for active safety functions such as Autonomous Emergency Braking (AEB) have traditionally been developed with the assumption that the driver is out of the loop, for example visually unaware of a conflict ahead or impaired. While this approach is understandable from a safety point of view, it may also result in warnings and interventions that are unnecessarily early for an attentive and non-impaired driver, therefore compromising the difficult balance between crash avoidance potential and driver acceptance (Ljung Aust, 2023). The fitment of Driver Monitoring systems creates the opportunity to incorporate the state of the driver into the warning and intervention sensitivity of active safety functions, contributing to a more appropriate response strategy. In Euro NCAP, this concept is referred to as Driver State Link (DSL).

Euro NCAP started promoting DSL from 2026, with score incentives both directly via the Driver Monitoring protocol and indirectly via the Crash Avoidance protocol. For 2029, Euro NCAP aims to go a step further by encouraging active safety functions to warn strictly when necessary, allowing suppression or delay of interventions in selected non-time-critical situations where the driver is deemed to be aware of the conflict ahead, while also incentivizing more effective interventions in severe impairment cases. In simple terms, the long-term objective is that the vehicle should only warn and intervene when the driver actually needs it.

.

2.3 Driver inattention

The transient driver states as part of the 2026 driver monitoring protocol are presented as a binary visual condition through which the driver is either looking at the forward roadway, or glancing off the road towards one of the specified gaze locations. Although this basic representation of visual attention does not present an issue for off-road glances, the “eyes on the road” state does not necessarily mean that the driver is visually aware of the conflict ahead, either due to cognitive distraction or simply due to the “looked but failed

to see" or "failed to look properly" cases (Department for Transport, 2014).

For that reason, the 2029 assessment aims to redefine driver inattention by also covering, at least partially, cases in which the driver is looking towards the forward roadway but is not aware of a conflict developing outside his or her effective visual scope. The first intention is not to solve every possible case, but to identify opportunities in simple and straightforward conflicts. One example would be a driver who is visually focused on following a lead vehicle but is not scanning a crossing pedestrian approaching from the nearside. Similar thinking is also relevant for Lane Support Systems, where visual attention and vehicle control cues may support suppression of unnecessary interventions, provided the driver is not impaired and is effectively managing the situation.

In this sense, the proposed framework moves beyond a purely binary distraction model and towards a more dynamic representation of risk that brings together driver visual attention and external events. This direction is also aligned with wider work in the field arguing for a broader attention-centric approach, where the question is not simply whether the driver is looking forward, but whether the driver is paying sufficient attention for the current situation and is ready to respond appropriately (Reimer, 2024).

The goal is therefore twofold: to ensure driver support when needed, and to avoid it when irrelevant.

***Figure 1**. Driver attention zone map concept for a potential conflict ahead (illustrative only).*

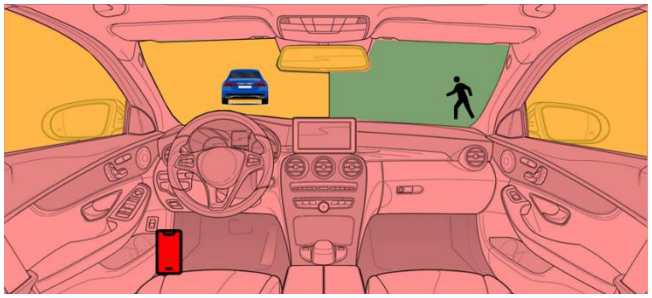


***Table 1** Zone definition and warning/intervention policy*

Zones	Intervention	Warning
Eyes ON conflict	Delay/suppression (e.g., FCW 1.2s instead of 1.7s)	N/A
Driving/ Monitoring	Contextual to support driver (e.g., unintentional lane change, crossing pedestrian not scanned)	N/A
Eyes OFF conflict	Timely to support visually distracted driver	N/A
Phone Use	Lockout of assisted features after repeated behaviour (e.g., ACC+LC)	VATS

2.4 Driver Monitoring accuracy and failsafe strategies

Although the benefits of DSL are evident, it is important to acknowledge that Driver Monitoring Systems are not all built to the same standard, and even state-of-the-art implementations remain subject to detection inaccuracy or system failure. In the worst case, this can lead to false negatives. For that reason, it is fundamental to define clear vehicle-response policies for situations where the driver state cannot be fully determined, for example where only head pose can be estimated, or where the system is unavailable altogether, for example due to sensor blockage. In such cases, the fallback strategy should ensure that the vehicle responds as if no Driver Monitoring information were available, in other words: do what you would do without DMS.

Regarding DMS performance, Euro NCAP introduced from 2026 a prerequisite linked to minimum Driver Monitoring performance in order for OEMs to benefit from the DSL concept. For 2029, the methodology is expected to evolve further so that the prerequisite is based more directly on minimum DMS accuracy rather than on the resulting Driver Monitoring score.

3. Assessment

For 2029, Euro NCAP aims to introduce an assessment methodology for Driver Monitoring that is built around three main pillars: Spot Checks, On-road Evaluation and Synthetic Data.

3.1 Spot Checks

Spot Checks on proving grounds have been the main verification method for the Euro NCAP DMS assessment since 2023. In an environmentally controlled setting, a test subject drives the assessed vehicle and executes a set of scripted behaviours to replicate selected transient and non-transient cases. Vehicle warning and intervention performance are then measured against the acceptance criteria. The main benefit of Spot Checks is repeatability and reproducibility in a low-risk setting. The main limitation is that, due to the resources involved, only a small number of cases can be checked, resulting in limited coverage of the full DMS performance envelope.

3.2 On-road Evaluation

From 2026, every vehicle assessed by Euro NCAP undergoes a naturalistic drive on public roads as part of the On-road Evaluation. The main purpose is to understand whether active safety technologies are fit for purpose in actual driving conditions, particularly from a driver acceptance perspective. Although the only element directly being assessed during the On-road Evaluation from 2026 to 2028 is the accuracy of SLIF (Speed Limit Information Function), Euro NCAP aims to extend the role of On-road Evaluation for other functions e.g. ACC, DMS, LSS. In the context of in-cabin functions, distraction warnings and associated interventions present a practical case to identify false positives during drive sessions, for example where the driver is confirmed by a ground-truth system to be looking at the forward roadway but a distraction warning is still triggered. This method can provide valuable insight into whether a DMS has been designed only to pass true positive cases, or

whether it has been developed with a genuinely human-centric approach.

3.3 Synthetic data

Simulation is a key pillar in the development and validation of engineered products, and synthetic data has shown potential to support the training and evaluation of machine-learning-based DMS. Euro NCAP has initiated a proof of concept to investigate whether synthetic data can also contribute to assessment. The objective is to expand verification coverage across demographic and noise-factor combinations that are difficult to capture efficiently through physical testing alone, therefore complementing Spot Checks rather than replacing them.

4. Discussion and conclusions

Before adopting the proposed criteria and assessment methodologies for future DMS assessments, close collaboration and technical alignment with industry will be fundamental to ensure that the expected societal benefits are realistically achievable in production vehicles. Multiple iterations are expected as the framework develops.

Overall, the 2029 direction presented here aims to move Driver Monitoring assessment beyond a mainly binary interpretation of visual distraction and towards a more context-aware framework for mitigating driver inattention risks. The proposal combines three main ideas: a broader interpretation of driver inattention, a more meaningful and safe use of Driver State Link, and a three-pillar assessment approach intended to improve both robustness and coverage. The overarching goal remains the same: to encourage fitment of active safety systems that support the driver effectively when support is needed, while avoiding unnecessary or poorly timed interventions that undermine driver acceptance.

5. Acknowledgments

The author would like to thank all the Euro NCAP Driver Engagement WG members for the invaluable expertise and safety-centric attitude.

References

Department for Transport. (2014). *Reported Road Casualties Great Britain: 2014 Annual Report.*

IET Digital Library. (s.f.). *Author Guide - IET Research Journals*.

Ljung Aust, M. (2023). *Key Human Factors Development Principles For Dms Enhanced Collision Avoidance System Development – ESV 23-0263.*

Reimer. (2024). *Supporting Driver Attention Holistically.* - DDI 2024

Exogenous Disruptions of Visual Attention and Gaze Abnormality

Huy Phan[1], Rafael Gonçalves[2*], Jorge Pardo[1], Brook Shiferaw[3], Yee Thung Lee[2], Jonny Kuo[3], Mike Lenné[3], Natasha Merat[2], Mohammed Elhenawy[1], Sebastien Glaser[1], Ashish Bhaskar[1], Ronald Schroeter[1]

[1] Queensland University of Technology, Brisbane, Queensland, Australia
[2] University of Leeds, Leeds, UK
[3] Seeing Machines, Melbourne, Victoria, Australia
[*] r.goncalves@leeds.ac.uk

Abstract: Drivers' gaze is guided by an interaction between endogenous attention and exogenous environmental saliences. Disruptions may indicate inadequate attention allocation and increased safety risk. Yet, defining expected gaze remains challenging due to contextual complexity and strong inter-individual variability. We present a self-supervised framework for detecting gaze abnormalities using eye-tracking data collected in a driving simulator. Gaze is encoded using a contrastive learning objective that distinguishes gaze patterns during free lane driving from disrupted patterns by a hazardous jaywalking pedestrian. The framework explicitly models inter-individual variability by learning participant-specific representations from diverse gaze. By mapping gaze sequences into a gaze embedding space, the framework categorises patterns as expected or deviated without extensive labelling. The methodology tests the feasibility of adaptive gaze modelling, laying the groundwork for human-centred driver monitoring systems capable of identifying critical visual attention shifts with greater sensitivity than traditional models.

1. Introduction

Drivers' visual scanning patterns are a product of the interaction between endogenous modulations of attention and exogenous saliences from the external environment (Carrasco, 2011; Trick & Enns, 2009). Studies on visual attention have suggested that because environmental context and specific task demands largely dictate expected visual activity, gaze remains highly situational (Hayhoe & Ballard, 2009). In that sense, several studies have sought to identify general characteristics that constitute drivers' expected visual scanning patterns across a wide range of driving scenarios (see e.g., Salvucci & Liu, 2002; Tijerina et al., 2005; Lenné, 2024; Wilkie et al., 2010; Itkonen et al., 2015). Despite potential context and individual-driven variations, saliency map models suggest that a situation's unique visual demands and expected patterns can be effectively modelled and predicted (Borji & Itti, 2013).

Inconsistencies between scenario demands and observed gaze patterns provide a potential measurement for driver monitoring systems to estimate a driver's state. Disruptions to these expected patterns can generally be attributed to endogenous modulations driven by the driver's state or changes in external (exogenous) environmental saliences, diverting drivers' visual attention away from task-related demands. For instance, alcohol intoxication is linked to higher gaze entropy (Shiferaw et al., 2019b). As for exogenous disruptions, an information-dense user interface increases drivers' dwell time away from the road centre (Gonçalves et al., 2022). However, defining a universal expectation of gaze patterns is extremely complex, as it is heavily influenced by both environmental context and individual participant variability (Goodridge et al., 2024).

Gaze patterns reflecting attentional behaviour can be effectively captured using mathematical models, with neural networks offering significantly higher scalability for complex datasets. Although various modelling approaches have been attempted to quantify these patterns (Shiferaw et al., 2019a, Pardo et al., 2025; Gonçalves et al., 2019; Underwood et al., 2003), neural networks have demonstrated high accuracy in generating gaze predictions (Helwan et al., 2022; Ngo and Manjunath, 2017).

Self-supervised neural networks can learn spatial and temporal structures directly from raw gaze pattern time-series, without prior feature engineering (Pathirana et al., 2022). Building on this, anomaly detection defines probabilistic deviations from expected gaze (Johari et al., 2025). We can define an abnormality as a deviation in gaze under specific spatio-temporal conditions, relative to prior observations from the same or similar drivers (Phan et al., 2025). Assuming training data adequately represents the variation space for a given task/context, detected abnormalities can be causally attributed to disruptions in driver state or the presence of saliencies.

Despite the framework's theoretical plausibility, the hypothesis lacks empirical supporting evidence. A validation study needs to be conducted, where the capability of the machine learning tool to detect evident abnormalities in experimentally controlled comparable situations is used as a benchmark for the theoretical validity of the construct. In this work, we investigate the following research question: *How do expected and deviated gaze patterns manifest across different drivers when exposed to exogenous disruptions?*

2. Method

To address the research question, we will develop a self-supervised gaze abnormality detection algorithm using driving simulator experimental data. The study is assessing whether the observed abnormalities are caused by the

presence of hazardous situations (characterised by a jaywalking pedestrian) when compared to the baseline data (free lane).

Our methodology converts time-series gaze data into a quantifiable metric of gaze abnormality. We base this approach on the principle that driving context dictates a baseline of expected visual behaviour. When a hazard enters the environment, it draws the driver's attention away from the baseline, generating a potential abnormality. Data is fed into a self-supervised model to learn a compact representation of the driver's natural scanning patterns. By utilising a contrastive learning objective (see Hjelm et al., 2019; Khosla et al., 2020; Chen et al., 2020), the framework learns to distinguish between these normative patterns and the disrupted patterns caused by jaywalking pedestrians.

2.1 Dataset

The scenarios are extracted from a motion-based, high-fidelity driving simulator experiment (see Lee et al., 2024). The experiment consists of 44 participants (24 males and 20 females), aged between 25 and 63 years (M = 40.23, SD = 11.19). The eye-tracking system we are using for this study is Tobii Glasses 3 at 50 Hz.

The experiment simulated an urban UK scenario in which drivers drove a vehicle at 30 miles per hour on a straight, unobstructed road. The dataset is divided into two types of road segments, each lasting approximately 8 seconds (~100 m), designed for this analysis. The first type consists of baseline segments representing free lane driving in a sparse road environment, with no leading vehicles or hazards present. These segments serve as reference blocks for characterising normal gaze. Each participant experienced seven baseline segments. The second type consists of a hazardous event segment of similar length, in which a pedestrian suddenly initiates a jaywalking manoeuvre by entering the roadway without prior signalling or waiting behaviour. This event segment is used to examine deviations from expected gaze. Two event types are included: a left pedestrian crossing from a side alley at 0.75 m/s (TTC ≈ 2 s) and a right pedestrian crossing at 1.5 m/s (TTC ≈ 4 s).

2.2 Modelling

In order to capture the nuances of driver attention, the data labelling process transforms raw gaze coordinates into spatio-temporal sequences, designed for the neural networks to extract multi-scale features. This strategy implements a dual-track temporal granularity: global sequences, which encompass the entirety of a segment to model macroscopic gaze dispersion, and local sequences, which are derived from velocity-based ocular event parsing (e.g., saccades and gaze fixations; Salvucci & Goldberg, 2000). By labelling data at these different scales, the model can simultaneously learn high-level behavioural trends and the fine-grained characteristics of gaze.

The representation component follows the self-supervised learning framework to model gaze time series and learn meaningful spatio-temporal representations directly from unlabeled gaze data. By using different types of time-series anomaly detection models (Darban et al., 2024), the framework transforms these sequences into informative spatio-temporal embeddings (Karessli et al., 2017), which function as vector representations synthesised from either raw gaze samples, detected ocular events, and spatial dispersion.

Our novel contrastive component extends reconstruction learning and operates as a model-agnostic approach. It enables time-series representation models to statistically identify the largest separation between expected and deviated groups (whose distributions align with gaze abnormality theory) in the gaze embedding space by generating multiple binary partitioning iterations that capture different semantic properties of gaze. Each iteration comes with a reconstruction score that acts as a proxy for interpreting deviated patterns.

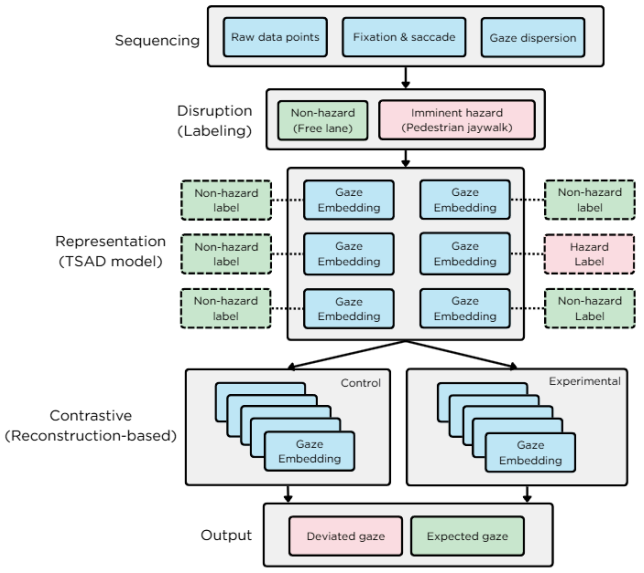


Fig 1. *Model's architecture.*

2.3 Validation

The model validation will be conducted as follows: For each participant, the segments will be divided into two groups. The control group, as a subset of five out of the seven baseline segments, and the experimental group, including the two hazardous event segments, together with the remaining two baseline segments. All possible combinations of training and evaluation sets will be generated. Discrepancy scores will be computed for each evaluation set, and the evaluation sets will be ranked by these scores. The top pairs of segments with the highest discrepancy scores for each group are isolated for a comparative analysis.

For the experimental group, hazardous event segments are expected to appear among these top pairs, whereas for the control group, discrepancy scores are expected to remain low. Finally, inter-driver pattern analysis will be performed by comparing evaluation pairs across drivers. The effectiveness of the models will be validated by calculating the percentage of cases correctly labelled (i.e., hazardous events as "deviated patterns"). A statistical test

will be applied to estimate differences between expected and deviated patterns.

3. Conclusions

This work presents a framework for investigating gaze abnormality as a marker of exogenous disruptions in visual attention during manual driving. Using a self-supervised, contrastive learning architecture, we will assess gaze deviations in driving simulation data. A key expected contribution is a gaze abnormality algorithm that models visual deviations conditioned on hazard disruption labels. The framework lays the groundwork for applications in participant-specific driver monitoring systems. The data is currently being analysed. The expected results will assess the validity of the proposed framework and provide an understanding of how deviations in gaze patterns can inform the development of next-generation active safety tools.

4. Acknowledgments

This research was supported by the Australian Government through the Department of Education's National Industry PhD Program (project 36293) and Seeing Machines through the Chair in Empathic Machines project, including financial aid and research resources. The views expressed herein are those of the authors and are not necessarily those of the Australian Government or the Department of Education.

Data for this study were supported by the EPSRC HAROLD project (HAzards, Road Lighting and Driving, EP/S003576/1), with additional funding provided through the UKRI IAA Grant Scheme (IAA4176).

References

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition, 17*(6–7), 1185–1204. https://doi.org/10.1080/13506280902978477

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 185–207. https://doi.org/10.1109/TPAMI.2012.89

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human Factors, 59*(5), 734–764. https://doi.org/10.1177/0018720817690639

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999–1010. https://doi.org/10.3758/BF03194829

Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., Lee, Y. M., Madigan, R., & Merat, N. (2024). Gaze entropy metrics for mental workload estimation are heterogeneous during hands-off level 2 automation. *Accident Analysis & Prevention, 202*, 107560. https://doi.org/10.1016/j.aap.2024.107560

Helwan, A., Ma'aitah, M. K. S., Uzelaltinbulat, S., Altobel, M. Z., & Darwish, M. (2022). Gaze prediction based on convolutional neural network. In M. Al-Emran, M. A. Al-Sharafi, M. N. Al-Kabi, & K. Shaalan (Eds.), *Proceedings of the International Conference on Emerging Technologies and Intelligent Systems* (pp. 215–224). Springer. https://doi.org/10.1007/978-3-030-85990-9_1 8

Itkonen, T., Pekkanen, J., & Lappi, O. (2015). Driver gaze behavior is different in normal curve driving and when looking at the tangent point. *PLOS ONE, 10*(8), e0135505. https://doi.org/10.1371/journal.pone.0135505

Johari, K., Kim, J. J., Yow, W. Q., & Tan, U. X. (2025). A foundation model-based framework for unsupervised gaze anomaly detection. *Knowledge-Based Systems, 324*, 113774. https://doi.org/10.1016/j.knosys.2025.113774

Kastovsky, D. (1992). Semantics and vocabulary. In R. M. Hogg (Ed.), *The Cambridge history of the English language: Vol. 1. The beginnings to 1066* (pp. 290–408). Cambridge University Press.

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies, 76*, 35–50. https://doi.org/10.1016/j.trc.2017.01.001

Ngo, T., & Manjunath, B. (2017). Saccade gaze prediction using a recurrent neural network. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 3435–3439). IEEE. https://doi.org/10.1109/ICIP.2017.8296920

Pardo, J., Gonçalves, R. C., Li, X., Kuo, J., Yang, S., Schroeter, R., Merat, N., & Lenné, M. (2025). Using gaze to understand shifts in visual attention during conditional automation with non-driving related tasks. In S. Viller et al. (Eds.), *Proceedings of the 36th Australasian Conference on Human-Computer Interaction (OzCHI '24)* (pp. 395–407). ACM. https://doi.org/10.1145/3726986.3727030

Pathirana, P., Senarath, S., Meedeniya, D., & Jayarathna, S. (2022). Eye gaze estimation: A survey on deep learning-based approaches. *Expert Systems with Applications, 199*, 116894. https://doi.org/10.1016/j.eswa.2022.116894

Petrică, A. C., Stancu, S., & Tindeche, A. (2016). Limitation of ARIMA models in financial and monetary economics. *Theoretical and Applied Economics, 23*(2), 19–42.

Phan, H. T., Yu, Y., Gonçalves, R. C., Kuo, J., Glaser, S., Elhenawy, M. M. Z., Schroeter, R., & Bhaskar, A. (2025). Self-supervised learning for detecting local contextual anomalous gaze patterns of individual drivers in level 3 automated driving. In *Adjunct Proceedings of the 17th*

International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 145–150). ACM. https://doi.org/10.1145/3744335.3758496

Salvucci, D. D., & Liu, A. (2002). The time course of a lane change: Driver control and eye-movement behavior. *Transportation Research Part F: Traffic Psychology and Behaviour, 5*(2), 123–132. https://doi.org/10.1016/S1369-8478(02)00011-6

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews, 96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007

Shiferaw, B. A., Crewther, D. P., & Downey, L. A. (2019). Gaze entropy measures detect alcohol-induced driver impairment. *Drug and Alcohol Dependence, 204*, 107519. https://doi.org/10.1016/j.drugalcdep.2019.06.021

Tan, Z., Wang, M., Xie, J., Chen, Y., & Shi, X. (2018). Deep semantic role labeling with self-attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.

Trick, L. M., & Enns, J. T. (2009). A two-dimensional framework for understanding the role of attentional selection in driving. *Proceedings of the Annual Meeting of the Cognitive Science Society*. https://api.semanticscholar.org/CorpusID:38688206

Tijerina, L., Garrott, W. R., Stoltzfus, D., & Parmer, E. (2005). Eye glance behavior of van and passenger car drivers during lane change decision phase. *Transportation Research Record, 1937*(1), 37–43. https://doi.org/10.1177/0361198105193700105

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., & Krishnan, D. (2020). Supervised contrastive learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 18661–18673). Curran Associates, Inc.

Hjelm, R. D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., & Bengio, Y. (2019). Learning deep representations by mutual information estimation and maximization. *arXiv*. https://arxiv.org/abs/1808.06670

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *arXiv*. https://arxiv.org/abs/2002.05709

Lee, Y. T., Gonçalves, R. C., Ozturk, I., Proctor, K., & Merat, N. (2024). Using a driving simulator study to evaluate drivers' hazard avoidance under different automation levels during day- and night-time conditions. *Abstract Book of the 9th International Conference on Driver Distraction and Inattention* (pp. 42–44). University of Michigan Transportation Research Institute.

Kircher, K., & Ahlström, C. (2017). Minimum required attention: A human-centered approach to driver inattention. *Human Factors, 59*(3), 471–484. https://doi.org/10.1177/0018720816676473

Gonçalves, R. C., Louw, T. L., Madigan, R., Quaresma, M., Romano, R., & Merat, N. (2022). The effect of information from dash-based human–machine interfaces on drivers' gaze patterns and lane-change manoeuvres after conditionally automated driving. *Accident Analysis & Prevention, 174*, 106726. https://doi.org/10.1016/j.aap.2022.106726

Gonçalves, R., Louw, T., Madigan, R., & Merat, N. (2019, June)**.** Using Markov chains to understand the sequence of drivers' gaze transitions during lane-changes in automated driving. In *Proceedings of the 10th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. University of Iowa.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics, 46*(6), 629–646. https://doi.org/10.1080/0014013031000103682

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). *Deep contextualized word representations*. arXiv. https://arxiv.org/abs/1802.05365

Le-Khac, P. H., Healy, G., & Smeaton, A. F. (2020). Contrastive representation learning: A framework and review. *IEEE Access, 8*, 193907–193934. https://doi.org/10.1109/ACCESS.2020.3031549

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv. https://arxiv.org/abs/1810.04805

Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C. C., & Salehi, M. (2024). Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys, 57*(1), Article 15, 1–42. https://doi.org/10.1145/3691338

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (pp. 71–78). Association for Computing Machinery. https://doi.org/10.1145/355017.355028

Wilkie, R. M., Kountouriotis, G. K., Merat, N., & Wann, J. P. (2010). Using vision to control locomotion: Looking where you want to go. *Experimental Brain Research, 204*(4), 539–547. https://doi.org/10.1007/s00221-010-2321-4

Lenné, M. G., Fitzharris, M., Fredriksson, R., Palao, A., & Young, M. S. (2024, October). Occupant status monitoring: A framework to integrate human factors, technology & policy to realise injury reductions across the safety chain, *International Conference on Driver Distraction and Inattention (DDI 2024)*.

Karessli, N., Akata, Z., Schiele, B., & Bulling, A. (2017). Gaze embeddings for zero-shot image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6412–6421). https://doi.org/10.1109/CVPR.2017.679

Disentangling alcohol and sleepiness effects on driver monitoring system metrics

Lovisa Andersson [1], Emmy Lindgren [1], Raimondas Zemblys [2], Christer Ahlström [1,3*]

[1] Linköping University, Dept of Biomedical Engineering, Linköping, Sweden
[2] Smart Eye AB, Gothenburg, Sweden
[3] Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden
[*] christer.ahlstrom@vti.se

Abstract: Camera-based driver-monitoring systems (DMS) are increasingly proposed for detecting alcohol-related impairment, even though the behavioural metrics they rely on are influenced not only by alcohol but also by other factors such as sleepiness. As a consequence, DMS tend to perform better on the (sedative) descending than the (stimulant) ascending limb of the BAC curve. The present study aims to disentangle the contributions of BAC, sleepiness, and intoxication phase to ocular behaviour. Seventy-six drivers completed test-track drives at varying intoxication and sleepiness levels across ascending and descending BAC phases. Ocular metrics were extracted from in-vehicle cameras. Mixed-effects models assessed the effects of BAC, subjective sleepiness, and BAC phase. A matched-pair analysis (BAC-matched and BAC + KSS-matched) quantified phase-dependent differences using Cliff's delta. The results show that many ocular features, especially blink duration, fixation duration, PERCLOS, and saccadic velocity, varied with both BAC and sleepiness. Our findings indicate that the asymmetry between ascending and descending phases in ocular behaviour is largely explained by increased sleepiness on the descending limb. This pattern reflects alcohol's sedative action during elimination, which elevates sleepiness and thereby affects the ocular metrics that DMS rely on. Thus, the phase differences do not reflect a direct oculomotor effect of BAC trajectory, but an indirect effect mediated through sleepiness. For safety applications, observable impairment rather than BAC may still be the most appropriate detection target, reinforcing the need for continuous real-time DMS.

1. Introduction

There is increasing interest in extending the capabilities of driver-monitoring systems (DMS) beyond their original focus on detecting sleepiness and distraction by using the same camera-based sensors to identify alcohol-related impairment. Camera-based features reflect behavioural changes rather than blood alcohol concentration (BAC) directly, meaning their specificity to alcohol (versus sleepiness or distraction) is limited without additional data, and moreover, their ability to detect impairment tends to be strongest at higher alcohol levels (Ahlström et al., 2025; Keshtkaran et al., 2024; Prendez et al., 2024).

Evidence from controlled simulator studies shows that DMS-derived ocular indicators are more sensitive during the descending limb of the BAC curve than during the ascending limb, meaning that observable impairment appears greater even when BAC levels are comparable (Shiferaw et al., 2025). However, a systematic review by Holland and Ferner (2017) demonstrated the opposite pattern for subjective intoxication: at identical BAC, people report feeling less intoxicated (29 % lower) and show a markedly increased willingness to drive (207 % higher) on the descending limb, despite substantially worse objective performance (96 % deterioration) and impaired inhibitory control (30 % worse). This asymmetry between subjective and objective impairment is a hallmark of the Mellanby effect (Mellanby, 1919), in which subjective impairment is greater on the ascending limb even though behavioural impairment may persist or intensify on the descending limb.

Physiological and behavioural indicators also follow phase-dependent trajectories. Eye-movement studies reveal diverse patterns: gaze-entropy measures remain impaired during both ascending and descending phases, fixation metrics show primarily ascending-phase impairment, and saccadic latencies follow biphasic dynamics reflecting early stimulant and later sedative pharmacological effects (Jäntti et al., 2004; Shiferaw et al., 2019).

Prior research has also shown that sleepiness is typically higher on the descending limb, even when BAC levels are matched (Shiferaw et al., 2025). Because sleepiness influences many of the same ocular indicators used for alcohol detection, the two states may confound one another unless explicitly disentangled. The objective of this study is to examine how different ocular features vary with BAC and sleepiness. Using mixed-effects regression and matched-pair methods designed to isolate phase-dependent effects, we evaluate the relative contributions of BAC, sleepiness, and BAC trajectory phase to ocular behaviour. The study seeks to answer the following questions:

1. Which ocular features vary systematically with BAC?
2. Which features vary with subjective sleepiness, as measured by the Karolinska Sleepiness Scale (KSS)?
3. Do ocular features differ in ascending and descending alcohol phases at comparable BAC levels, and to what extent are these differences attributable to sleepiness?

2. Method

Data were collected in a controlled test-track study designed to capture ocular behaviour across a wide range of BAC values, subjective sleepiness levels, and both ascending and descending phases of alcohol intoxication. Seventy-six licensed drivers participated. Four experimental schedules were used to systematically vary intoxication trajectories and time of day. Two daytime groups drove while alert, whereas two night-time groups completed drives where sleep pressure

Table 1. *Mixed-effects regression results for all ocular features. Directions of change are based on relevant main-effect terms, with significance stars showing Bonferroni adjusted p values (p < .05 = *, p < .01 = **, p < .001 = ***). Phase effects represent differences between ascending and descending BAC trajectories, accompanied by their corresponding phase coefficients in models without (M0) and with (M1) KSS adjustment, as well as the percentage attenuation of the phase effect after sleepiness adjustment.*

Feature	BAC effect	KSS effect	Phase effect	Phase M0 coefficient	Phase M1 coefficient	Attenuation
Blink Duration (ms)	↑***	↑	↑***	0.502***	0.347*	30.9%
Eyelid Opening Speed (m/s)	↓***	↓	↓***	-0.365***	-0.542***	-48.4%
Saccade Frequency (/min)	↓***	↓*	↓***	-0.780***	-0.514	34.1%
Eyelid Closing Speed (m/s)	↓***	↑	↓***	-0.104***	0.095	9.1%
PERCLOS (%)	↑***	↑*	↑***	0.772***	0.520	32.6%
GTE (bits)	↓***	↓	↓***	-0.741***	-1.100*	-48.4%
Saccade Mean Velocity (°/min)	↓***	↓	↓***	-0.139***	0.005	96.2%
Saccade Peak Velocity (°/min)	↓***	↓	↓***	-0.165***	-0.142	14.0%
Blink Frequency (/min)	↑	↑	↑**	0.368**	0.374	-1.8%
PRC (%)	↑	↑	↑***	0.185***	0.338**	-83.2%
SGE (bits)	↓	↓*	↓***	-0.470***	-0.623	-32.6%
SampEn (nats)	↑	↑*	↑**	0.162**	0.590*	-264.9%
Saccade Amplitude (°)	↓***	↑	↓***	-0.201***	-0.004	98.1%
Eyelid Closing Amplitude (cm)	↑*	↑**	↑***	0.109***	0.556***	-411.8%
Fixation Duration (ms)	↑***	↑*	↑***	0.453***	0.647*	-42.9%
Eyelid Opening Amplitude (cm)	↑	↑*	↑**	0.077**	0.489***	-536.0%
Time Lag (s)	↑	↑	↑	0.109	1.362	-1149.6%
Main Sequence Slope (°s/m)	↑	↑	↑	0.025	0.137	-440.2%

increased naturally. Depending on group allocation, participants completed between three and six driving sessions with target BAC levels between 0 and 1.2‰, ensuring coverage of both increasing and decreasing alcohol concentrations. More information about the protocol can be found in Ahlström et al. (2025).

Before each session, BAC was measured using a calibrated breathalyser, and participants rated their sleepiness using the nine-point Karolinska Sleepiness Scale (KSS). Additional KSS ratings were provided every fifth minute during each approximately 25-minute drive. Eye-tracking data were derived at 60 Hz using a three-camera Smart Eye Pro 13.0.0 system. This protocol produced repeated measures spanning different levels of intoxication, alertness, and alcohol-trajectory phase, enabling analysis of BAC effects, sleepiness effects, and ascending–descending differences in ocular features.

Eighteen ocular features were extracted for each non-over-lapping 5-minute analysis window, matching the temporal resolution of the KSS ratings collected during the drives: Blink Duration (ms), Eyelid Closing Amplitude (cm), Eyelid Opening Amplitude (cm), eyelid Closing Speed (m/s), Eyelid Opening Speed (m/s), Fixation Duration (ms), Saccade Amplitude (°), Saccade Mean Velocity (°/min), Saccade Peak Velocity (°/min), Blink Frequency (/min), PERCLOS (%), Saccade Frequency (/min), Main Sequence Slope (°s/m), Percent Road Centre (PRC, %), Stationary Gaze Entropy (SGE, bits), Gaze Transition Entropy (GTE, bits), Sample Entropy (SampEn, nats), and Time Lag (s).

For each feature, two mixed-effects regression models were fitted: M0 with BAC × phase and M1 with BAC × phase plus KSS and their interactions. Spline bases with two degrees of freedom captured non-linear effects of BAC. Participant was included as a random effect to account for repeated measures and inter-individual variability. These models provided estimates of whether each feature was associated with alcohol intoxication, sleepiness, or both. The extent to which the phase effect was attributable to KSS was quantified as $(|\text{phase}_{M0}| - |\text{phase}_{M1}|)/|\text{phase}_{M0}|$.

To examine differences between ascending and descending BAC levels, within-participant matched-pair analyses were conducted using a tolerance of $|\Delta\text{BAC}| \leq 0.2‰$. A secondary, stricter matching additionally required $|\Delta\text{KSS}| \leq 1$ to control for differences in subjective sleepiness. Differences between matched pairs were quantified using Cliff's delta.

3. Results

Mixed-effects regression revealed consistent main effects of BAC, KSS, and phase (ascending vs. descending) on multiple ocular features (Table 1). Higher BAC was associated with longer blink durations, slower eyelid movements, and reduced saccade amplitude and velocity. These associations were robust across both M0 and M1, indicating a direct effect of alcohol on oculomotor behaviour. Sleepiness also showed strong main and/or interaction effects, including on blink duration, fixation duration, PERCLOS, and saccade dynamics, reflecting substantial overlap between alcohol- and sleep-related influences.

Phase differences were apparent across several features in M0, with many measures differing systematically between descending and ascending limbs at the same BAC. However, many of these effects attenuated when subjective sleepiness was included in M1, indicating that part of the observed phase gap reflects sleepiness rather than alcohol alone.

The matched-pair analysis quantified ascending–descending differences at comparable BAC (Table 2). Large

Table 2. *Cliff's delta effect sizes quantifying differences between ascending and descending BAC phases, when matching observations by BAC only versus both BAC & KSS. Positive values indicate higher feature values on the descending limb. Significance stars correspond to Bonferroni adjusted p values (p < .05 = *, p < .01 = **, p < .001 = ***). The rightmost column reports the difference between effect sizes, indicating the extent to which sleepiness contributes to the observed differences.*

Feature	δ_{BAC}	N_{BAC}	$\delta_{BAC\&KSS}$	$N_{BAC\&KSS}$	$\lvert\delta_{BAC\&KSS}\rvert$-$\lvert\delta_{BAC}\rvert$
Blink Duration (ms)	0.766***	111	0.585***	53	-0.181
Eyelid Opening Speed (m/s)	-0.568***	111	-0.472*	53	-0.096
Saccade Frequency (/min)	-0.517***	120	-0.236	55	-0.280
Eyelid Closing Speed (m/s)	-0.495***	111	-0.321	53	-0.175
PERCLOS (%)	0.473***	110	0.308	52	-0.165
GTE (bits)	-0.418***	110	-0.255	51	-0.163
Saccade Mean Velocity (°/min)	-0.417***	120	-0.164	55	-0.253
Saccade Peak Velocity (°/min)	-0.400***	120	-0.164	55	-0.236
Blink Frequency (/min)	0.369**	111	0.509**	53	0.140
PRC (%)	0.351**	111	0.154	52	-0.198
SGE (bits)	-0.351**	111	-0.154	52	-0.198
SampEn (nats)	0.333**	120	0.236	55	-0.097
Saccade Amplitude (°)	-0.333**	120	-0.164	55	-0.170
Eyelid Closing Amplitude (cm)	0.297*	111	0.434*	53	0.137
Fixation Duration (ms)	0.233	120	0.127	55	-0.106
Eyelid Opening Amplitude (cm)	0.225	111	0.283	53	0.058
Time Lag (s)	0.065	107	0.200	50	0.135
Main Sequence Slope (°s/m)	0.033	120	-0.018	55	-0.015

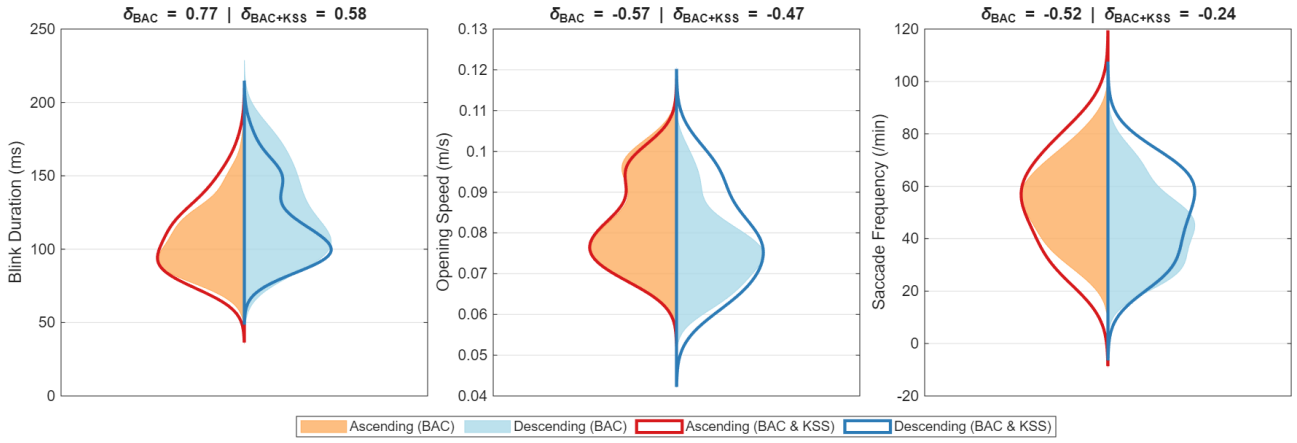


Fig. 1. *Split violin plots of the three features with largest effect size for BAC-matched pairs in ascending versus descending limbs. Filled distributions are from the pairs matched only on BAC. The outlines are from the pairs that are matched on BAC & KSS. Effect sizes are reported in the titles.*

phase differences were observed for several features when matching on BAC alone. Phase effects generally weakened when controlling for sleepiness. For example, Saccade Frequency decreased from δ_{BAC} = -0.517 to $\delta_{BAC\&KSS}$ = -0.236. Across features, both alcohol and sleepiness contributed to ascending–descending differences, albeit to varying degrees.

Fig. 1 shows distributions for the three features with the largest effect sizes between ascending and descending phases when matching observations on BAC only (filled distributions) versus BAC and KSS (outlined distributions). With BAC-only matching, blink durations were longer, opening speeds slower, and saccades less frequent on the descending phase. After controlling for KSS, the ascending and descending distributions became more similar, consistent with the attenuations observed in Table 2.

4. Limitations

Several limitations should be considered when interpreting the present findings. First, although the study design included both daytime and night-time driving sessions, sleepiness and circadian timing were not experimentally controlled, making it difficult to fully disentangle sleep-pressure effects from descending-phase alcohol effects. Second, sleepiness increased not only as a function of time of day but also within each drive, meaning that time-on-task may have contributed to some of the phase-related

differences. Third, although BAC matching ($|\Delta BAC| \leq 0.2‰$) reduced confounding, this tolerance may be too wide (especially at low BAC levels), introducing residual mismatch in impairment states. Finally, the study was conducted on a closed test track, and the ecological validity of the findings may be limited compared with real-world driving environments with more complex cognitive and situational demands.

5. Conclusions

Our findings indicate that the asymmetry between ascending and descending phases in ocular behaviour can be largely explained by increased sleepiness on the descending limb, rather than by a reversal in alcohol-related impairment. This interpretation is consistent with earlier work showing that subjective sleepiness is higher on the descending limb even when BAC is matched (Shiferaw et al., 2025) and reflects the known sedative effect of alcohol during the elimination phase (Holland & Ferner, 2017).

For DMS applications, these results suggest that observable behavioural impairment—rather than BAC—is the appropriate detection target, as the same BAC may correspond to very different functional states depending on sleep pressure and intoxication phase. Consequently, continuous, real-time monitoring of behavioural indicators is essential for accurate assessment of driver fitness, irrespective of the underlying cause of impairment.

6. Acknowledgments

This work was partly funded by the Swedish Strategic Vehicle Research and Innovation Programme (VINNOVA FFI; Grant No. 2025-04148).

References

Ahlström, C., Zemblys, R., Finér, S., & Anund, A. (2025). The Combined Effect of Alcohol and Sleepiness on Eye Tracking Features: Implications for Driver Monitoring Systems. 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Copenhagen, Denmark.

Holland, M. G., & Ferner, R. E. (2017). A systematic review of the evidence for acute tolerance to alcohol – the "Mellanby effect". *Clinical Toxicology*, *55*(6), 545-556. https://doi.org/10.1080/15563650.2017.1296576

Jäntti, V., Lang, A., Keskinen, E., Lehtinen, I., & Pakkanen, A. (2004). Acute effects of intravenously given alcohol on saccadic eye movements and subjective evaluations of intoxication. *Psychopharmacology*. https://doi.org/10.1007/BF00427822

Keshtkaran, E., von Berg, B., Regan, G., Suter, D., & Gilani, S. Z. (2024, 3-8 Jan. 2024). Estimating Blood Alcohol Level Through Facial Features for Driver Impairment Assessment. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV),

Mellanby, E. (1919). *Alcohol: its absorption into and disappearance from the blood under different conditions*. The University Press.

Prendez, D. M., Brown, J. L., Venkatraman, V., Textor, C., Parong, J., & Robinson, E. (2024). *Assessment of Driver Monitoring Systems for Alcohol Impairment Detection and Level 2 Automation*. Department of Transportation. National Highway Traffic Safety Administration. Office of Vehicle Safety Research. https://rosap.ntl.bts.gov/view/dot/77647

Shiferaw, B., Crewther, D., & Downey, L. (2019). Gaze Entropy Measures Reveal Alcohol-Induced Visual Scanning Impairment During Ascending and Descending Phases of Intoxication. *Journal of Studies on Alcohol and Drugs*. https://doi.org/10.15288/JSAD.2019.80.236

Shiferaw, B., Zhu, X., Atken, B., Kuo, J., Downey, L., Lenne, M., & Hayley, A. (2025). *Beyond BAC Threshold: Developing a real-time alcohol-induced impairment detection strategy* 24th International Council on Alcohol, Drugs and Traffic Safety (ICADTS) Triennial Conference, Alcobaca, Portugal.

Micro- and Macro-Steering Reversal Analysis for Personalized Drowsiness Detection in Heavy Commercial Vehicles

Sanjana Hassan Ananda Kumar [1,2*] , David Öst [2] , Mats Jonasson[1]

[1] Chalmers University of Technology, Department of Mechanics and Maritime Sciences, Göteborg, Sweden
[2] Volvo Group Trucks Technology, Safe Efficient Driving & Services, Göteborg, Sweden
[*] sanjanaha101@gmail.com

Abstract: This research was a master's thesis that investigated and developed a real-time driver drowsiness detection model based on steering behaviour in heavy commercial vehicles. A derived metric of steering, steering reversal rate (SRR), was analysed to examine the correlation between driver's steering behaviour and drowsiness, measured using the Karolinska Sleepiness Scale (KSS). It was first established that the highest correlation between KSS and micro-reversals was at a gap size 3° and with that of macro-reversals was at gap size 6°. Although both micro- and macro- reversals showed good correlation with drowsiness, only micro-reversals were used for model development because they exhibited a more pronounced change with increasing drowsiness and demonstrated a stronger correlation with KSS. This model defined a dynamic baseline for each driver based on their individual micro-reversal rate and any decline from this baseline was considered indicative of drowsiness, at speeds greater 65kmph (Council Regulation (EU), 2021/1341, 2021). This personalised model yielded a sensitivity of 58.8% and a false warning percentage of 2.5% while a generic model when applied to all drivers yielded a higher sensitivity of 72.1% but also a higher false warning percentage of 37.5%. (Sensitivity measured according to Council Regulation (EU), 2021/1341, 2021).

1. Introduction

Steering behavior has demonstrated considerable potential as an indicator of driver drowsiness (e.g., Altmüller, 2007, Gastaldi, et al., 2014, Thiffault, et al., 2003, Wang et al., 2016 & Zhang et al., 2016), as subtle variations in steering input reflect fluctuations in driver alertness. The steering behavior can be assessed in different ways and one of the more common ways is to analyze the derived steering metric called the steering reversal rate (SRR). SRR is defined as the number of changes in steering direction, per minute, between 1° to 10° angle, referred to as the gap size (Markkula & Engström, 2006).

Two of the prior studies (Altmüller, 2007 & Zhang et al., 2016) that have used steering reversal rate to investigate the correlation with driver drowsiness find an optimal gap size to be used for calculating steering reversals to conduct the analysis. However, Altmüller, (2007) arrives at the optimal value of gap size 6° with reference to PERCLOS* and video rating fatigue references, in a simulated environment and Zhang et al., (2016) arrives at the same value with reference to KSS but conducting an on-road experiment with 36 taxi drivers using the same passenger car. There is no previous literature to generalize this behavior across different vehicle segments and no optimal gap size identified separately for smaller and larger steering reversals to study driver drowsiness. Although Markkula & Engström, 2006 establish that 3° is the best gap size while measuring visual load component while performing secondary tasks, they do not include drowsiness in their analysis.

The study we conducted, however, was able to optimize the correlation with KSS across different gap sizes and identify that for micro-reversals, 3° is the most sensitive gap size and also corroborate the finding of Zhang et al. (2016) that 6° for large steering reversals is the most sensitive gap size with reference to KSS, even in commercial vehicles. Also, it was observed that the micro-reversals show a stronger correlation to KSS compared to that of macro-reversals.

Both micro- and macro-reversals derived at these gap sizes were further analyzed and used to develop a personalized model to detect a drowsy driver.

2. Method

2.1 Test preparation and data extraction

The log data used in this analysis were collected from drowsy driver tests conducted on heavy duty commercial vehicles in 2022 on a 5-kilometer-long ovular test track designed to replicate motorway driving conditions, as per the test methods outlined in Council Regulation (EU), 2021/1341, 2021.

The dataset comprised of 40 test drives conducted by 25 drivers, with no more than two test runs per individual. Of the 25 drivers, 20 were male and 5 were female. The drivers were internal Volvo drivers, and all had at least a Swedish C driver's license. The drivers were requested to be tired at the start of the test, by waking up earlier than normal and performing the test after a full shift, without consuming coffee, tea, energy drinks or similar.

Scale	Degree of Alertness/Sleepiness
1	Subject extremely alert
2	Very alert
3	Alert
4	Fairly alert
5	Neither alert nor in sleep mode
6	Few signs of sleepiness
7	Sleepy, no effort to keep alert
8	Sleepy, noticeable effort to keep alert
9	Extremely sleepy, great effort to keep alert, fighting with sleepiness

Fig. 1. *Karolinska Sleepiness Scale*

*PERCLOS = percentage of time pupils of eyes are 80% or more occluded over a specified time interval

The drivers were told to maintain the target speed of 70km/h manually. No type of cruise control was permitted during the test. There was a safety driver in the truck within the reach of the steering wheel and parking brake. The tests were ended if the driver fell asleep or rated him/herself as KSS 9 for 1 hour or if the safety driver had to intervene or at the latest after 3 hours of driving. Tests were also aborted if the driver or safety driver requested to do so, e.g. due to discomfort or if the safety could not be guaranteed.

The drivers along with the test leader were trained on the definitions of the KSS levels, before the test. During the test, the drivers rated themselves on the KSS every 10 minutes. A 10-minute interval was chosen to minimize driver distraction compared to the recommended 5-minute interval, while still maintaining adequate measurement accuracy.

The key signals extracted from the test data were the steering wheel angle (down sampled to 10Hz) and the KSS ratings (Åkerstedt & Gillberg, 1990) (refer Fig.1), provided by the drivers.

A linear relationship was assumed between the SRR and KSS ratings (Zhang et al., 2016 & Altmüller, 2007). The correlation was quantified using the Pearson correlation coefficient, a statistical measure of linear correlation. Fig 2(a) and (b) show the correlations between SRR and KSS ratings.

Analysis of the correlation across all gap sizes with KSS showed that the highest average correlation observed was -0.54 (p-value = 7.7×10^{-5}), corresponding to a gap size <3°, and +0.43 (p-value = 2.2×10^{-3}), corresponding to a gap size >6°. It was observed that the micro-reversals at 3° showed a stronger correlation compared to macro-reversals at 6°.

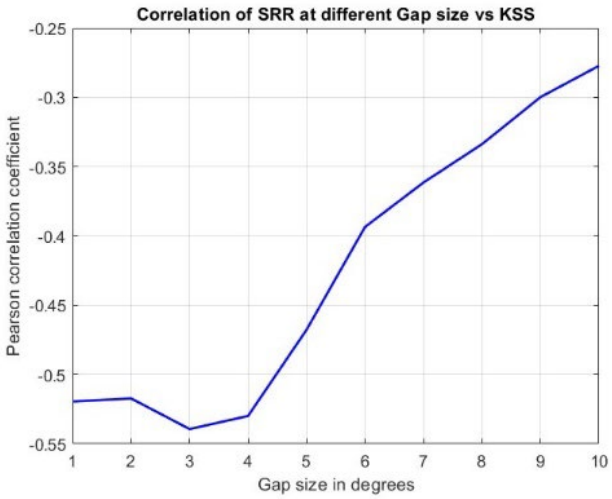
2.2 Model design and implementation

The model was designed based on the observation that micro-reversals decreased and simultaneously macro-reversals increased when the driver got drowsier.

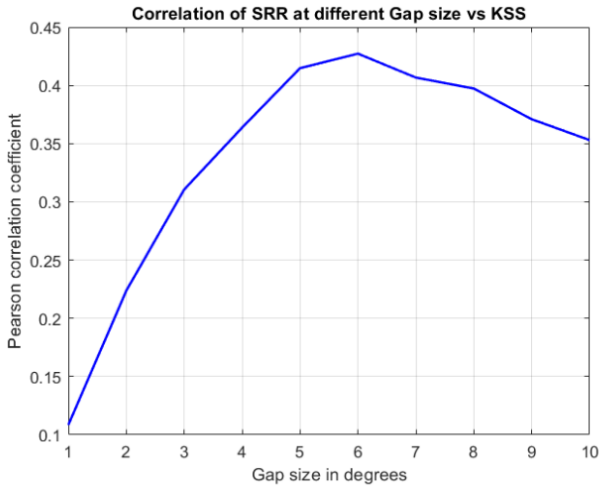
However, from Fig 3(a) and (b) it was evident that the change in macro-reversal values was not as distinct as that of micro-reversals especially between KSS 6 and KSS 8, although its trend in increase with KSS levels was visible. Overall, the micro-reversals showed a greater correlation to KSS when compared to macro-reversals. Hence, it seemed prudent to use only micro-reversals to develop the model.

The model set a baseline for each driver as the maxima of their micro-reversals because an increase in micro-reversals meant a decrease in driver drowsiness (Fairclough et al., 1999 & Ruijia et al., 2009). This baseline was dynamic in nature, i.e., if the driver got less drowsy over the course of the test, the baseline would increase accordingly. A drop in micro-reversals percentage from the baseline was computed continuously. Based on the drop percentage, the KSS values were calculated. If the KSS value was greater than 7 the model gave out a warning. These drop percentages were scaled according to the baseline, making the thresholds personalized to each driver. The scaling factor and in turn the drop percentages are the tuneable components of the model.

In contrast, another model with generic thresholds for all drivers was developed to compare the outcomes of a personalised vs a generic model for detecting driver drowsiness. This model utilized the micro-reversal distribution in Fig 3(a) to arrive at an absolute range of thresholds for each KSS level to assess where the driver's drowsiness level was.

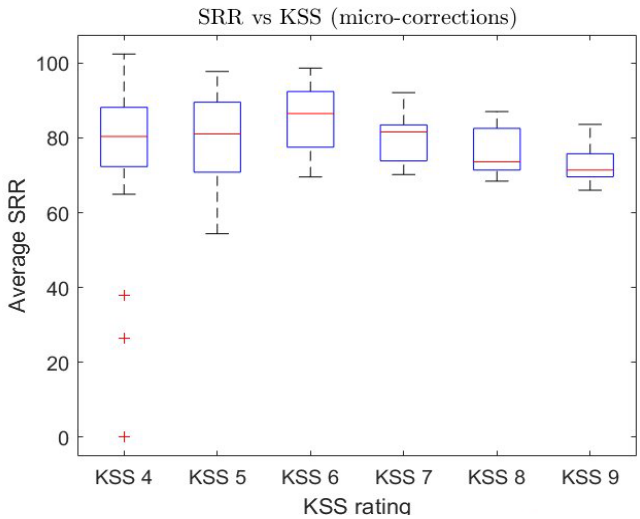


2 (a)

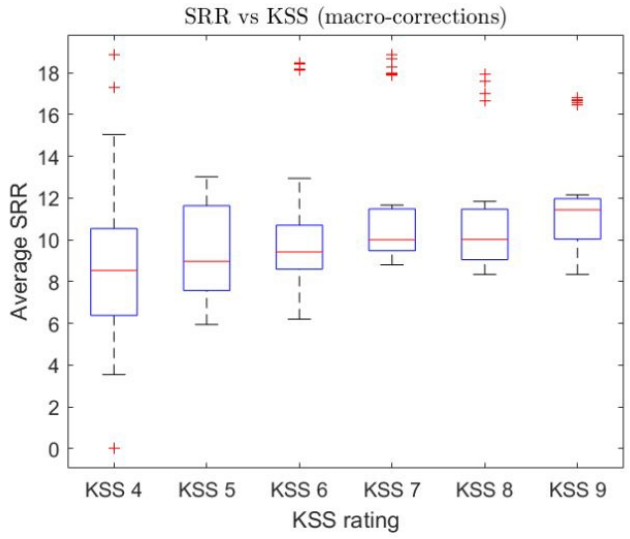


2 (b)

Fig. 2. *Pearson correlation coefficient between SRR & KSS* ***(a)*** *negative correlation – micro-reversals (gap size <1° to 10°),* ***(b)*** *positive correlation – macro-reversals (gap size > 1° to 10°)*



3 (a)



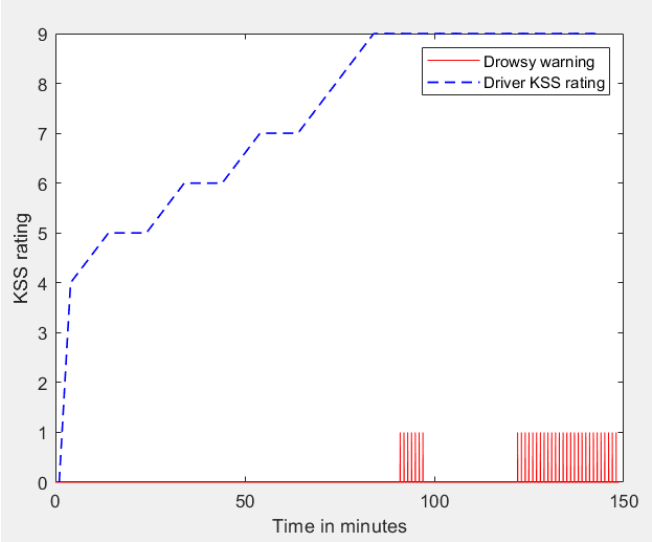
3 (b)

Fig. 3. *Steering reversal rate vs KSS rating -* ***(a)*** *micro-reversals at 3°* ***(b)*** *macro-reversals at 6°*

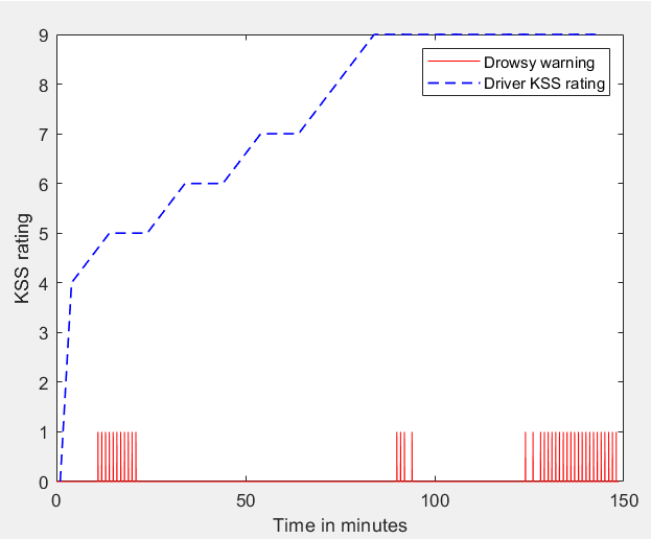
3. Results

The sensitivity of each model was calculated in accordance with the definitions outlined in clause 5 of Part 2 of Council Regulation (EU), 2021/1341, 2021. This method of calculating sensitivity varied slightly because it was based on the warning provided by each model when the driver identified himself/herself at KSS 7 or above and not if the model predicted the right KSS level of the driver. The personalised model had a sensitivity of 58.8%, and the generic model had a higher sensitivity of 72.1%, per drive. While the false warning percentage was also higher at 37.5% for the generic model and 2.5% for the personalised model, per drive. (Refer Fig 5 and Table 1).

The false warning percentage was calculated as the total number of false warnings divided by the total number of test drives. The sensitivity and false warning percentages can be varied according to the warning strategies adopted by the model.



5 (a)



5 (b)

Fig. 5. *Drowsiness detection and warning for a sample test drive -* ***(a)*** *Personalized model* ***(b)*** *Generic model*

in %	Personalized model	Generic model
Average sensitivity per driver	58,6	71,6
Average false warning per driver	4,0	60,0
Average sensitivity per drive	58,8	72,1
Average false warning per drive	2,5	37,5

Table 1. *Results - Sensitivity and false warning percentages*

4. Discussions

Will be presented at the conference.

5. Conclusions

Will be presented at the conference

References

Gastaldi, M., Rossi, R., and Gecchele, G. (2014). Effects of Driver Task-Related Fatigue on Driving Performance. Procedia—*Social and Behavioral Sciences, Vol. 111, 2014,* 955–964 https://doi.org/10.1016/j.sbspro.2014.01.130

Thiffault, P., Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention. 2003 May,35(3),* 381-91. https://doi.org/10.1016/s0001-4575(02)00014-3

Wang, X., Xu, C. (2016).Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accident Analysis & Prevention, 2016, Volume 95(B),* 350-57. https://doi.org/10.1016/j.aap.2015.09.002

Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, *52*(1–2), 29–37. https://doi.org/10.3109/00207459008994241

Zhang, H., Wu, C., Huang, Z., Yan, X., & Qiu, T. Z. (2016). Sensitivity of Lane Position and Steering Angle Measurements to Driver Fatigue. *Transportation Research Record: Journal of the Transportation*

Research Board, *2585*(1), 67-76. https://doi.org/10.3141/2585-08

Altmüller, T. (2007). *Driver Monitoring and Drowsiness Detection by Steering Signal Analysis*. (ISBN: 3832266011) [Doctoral dissertation, Universität der Bundeswehr München], Shaker publishers.

European Parliament and Council. (2021). *COMMISSION DELEGATED REGULATION (EU) 2021/1341 of 23 April 2021 laying down detailed rules concerning the specific test procedures and technical requirements for the type-approval of motor vehicles with regard to their driver drowsiness and attention warning systems.* EUR-Lex. http://data.europa.eu/eli/reg_del/2021/1341/oj

Markkula, G & Engström, J. (2006) *A Steering Wheel Reversal Rate Metric for Assessing Effects of Visual and Cognitive Secondary Task Load. Proceedings of the 13th ITS World Congress. 13th ITS World Congress*, London, UK. http://eprints.whiterose.ac.uk/131534/

Fairclough, SH., Graham, R. (1999) Impairment of driving caused by sleep deprivation or alcohol: a comparative study. *Human Factors. 1999*;*41(1),*118-28. https://doi.org/10.1518/001872099779577336

Ruijia F., Guangyuan Z., Bo C. (2009) An on-Board System For Detecting Driver Drowsiness Based on Multi-Sensor Data Fusion Using Dempster-Shafer Theory. *Proceedings of the International Conference on Networking, Sensing and Control; Okayama, Japan. March 2009*; 897–902. https://doi.org/10.1109/ICNSC.2009.4919399

Melman, T., Abbink, D. A., van Paassen, M. M., Boer, E. R., & de Winter, J. C. F. (2018). What determines drivers' speed? A replication of three behavioural adaptation experiments in a single driving simulator study. *Ergonomics*, *61*(7), 966–987. https://doi.org/10.1080/00140139.2018.1426790

Sun Y, Wu C, Zhang H, Zhou W, Li X, Zhang Q. (2021) Extraction of optimal fatigue-driving steering indicators considering individual differences. *IET Intell Transp Syst*. *2021*; *15*: 606–618. https://doi.org/10.1049/itr2.12048

Grahn, H., Ozturk, I. and Merat, N. (2024) How younger and older drivers' steering reversals change with cognitive distraction during both day and night-time driving. *Abstract Book, 9th International Conference on Driver Distraction and Inattention. International Conference on Driver Distraction and Inattention, 22-24 Oct 2024, Michigan, USA.* 38-41. https://ddi2024.org/wp-content/uploads/2024/10/DDI2024-ABSTRACTS.pdf

Driver attention during partially automated driving: Insights from reference drivers' glance behaviour

Claus Marberger [1*], Dimitri Löchner [1], Andreas Korthauer [1]

[1] Corporate Research, Robert Bosch GmbH, Robert-Bosch-Campus 1, Renningen, Germany
[*] claus.marberger@de.bosch.com

Abstract: Human inattention in partially automated driving is one of the key safety issues calling for effective countermeasures such as adequate driver monitoring. To better understand what constitutes "attention" in real world conditions, we conducted a study with reference drivers in various roles and across different road environments. Detailed inspections of their visual behaviour challenge common assumptions. These new insights can be used to design advanced methods for assessing driver engagement while using an L2 driving automation feature.

1. Introduction

Partial driving automation (SAE Level 2) requires attentive drivers capable of permanently recognizing potential hazards in the environment and staying ready to intervene when needed. At the same time, research has shown that human drivers may fail to accomplish this task sufficiently (Banks et al., 2018; Boos et al., 2020). To mitigate this undesirable effect, driver monitoring and intervention systems have been proposed (Mueller et al., 2021). However, setting the right criteria and thresholds to effectively warn a driver remains challenging.

First, although off-road glance durations of more than two seconds significantly increase the likelihood of being involved in a (near-)crash (Klauer et al., 2010), a fixed warning threshold of two seconds may generate a high number of nuisance alerts (Koniakowsky et al., 2023).

Second, legal or regulatory requirements on driver distraction mitigation (e.g., UNECE, 2023) focus on excessive distraction only (e.g., tolerating single off-road glance durations of up to 5 seconds), potentially missing many forms of safety-critical user behaviour.

Third, attention in driving is not only about setting thresholds for tolerating off-road glances but also ensuring that sufficient information is sampled from the environment to accomplish the task of driving (Kircher & Ahlstrom, 2017).

Fourth, research on driver inattention based on real-world multimodal driving data is scarce. While an integrated view of driver state, vehicle dynamics and traffic context has been proposed (Ahlström et al., 2021; Herbers et al., 2023), collecting such data in real driving settings can be cumbersome.

This work presents a study designed to investigate the glance behaviour of reference drivers under real-world driving conditions, focusing on three exemplary hypotheses:

- H1: The duration of long glances towards secondary task related areas decreases with vehicle speed
- H2: Glance behaviour towards mirrors is similar between supervised driving and manual driving
- H3: Traffic relevant objects in the frontal scene are fixated at least once while they are visible to the driver

2. Method

2.1 Sample

Data collection for this study intentionally focussed on reference drivers, represented by four skilled driving instructors recruited from local driving schools (two men and two women; minimum experience of supervising student drivers: ten years). We targeted driving instructors as study participants since they are highly trained to supervise an "imperfect driving system" and to intervene when required – exactly what is required and desirable from drivers in the context of L2 driving assistance.



Figure 1: *The reference track on public roads (including rural/urban road types and motorway)*

2.2 Vehicle and reference track

The research vehicle was based on a passenger car that has been modified to integrate additional controls for steering, accelerating/braking, and secondary driving controls. It was equipped with additional display devices and dedicated measurement systems for recording vehicle-related sensor data and data from (camera-based) interior monitoring. The setup enables the comprehensive simulation of automated driving features by a Wizard-of-Oz approach on test tracks and public roads (Marberger et al., 2019).

For this study, a team of two wizard operators were trained to simulate a SAE Level 2 hands-free feature (by a driving wizard) including realistic driver feedback (controlled by an HMI wizard) without drawing attention to their roles in the vehicle. The test participants could intervene in steering/braking at any time of the drive. The test track (see **Figure 1**) features multiple public road types (rural roads, motorway, urban roads) with various types of right-of-way regulations, and varying traffic densities. Completing the track took approximately 35-40 minutes.

2.2. Procedure and experimental conditions

All reference drivers were informed in advance about the study procedure and how data is processed within the project. Inside the vehicle, detailed instructions about the operation of the car and the L2 feature were given. The general task for all participants was to act during the drives as they deemed appropriate in their role as a responsible driver. After a familiarization phase each reference driver completed the drives under the following conditions in varying order:

- "max": Supervision of the L2 feature with unrestricted attentional resources (no secondary task)
- "SuRT": Supervision of the L2 feature with self-regulated engagement in a visual-manual secondary task on an installed tablet device (see ISO/TS 14198 and Yandel, 2016)
- "man": Manual driving with unrestricted attentional resources (no secondary task)

2.3 Data acquisition and pre-processing

Data collection included a combination of various sources of information about the driver state, vehicle dynamics, and the driving context. For the analyses of this publication, the following data has been acquired and/or generated:

- Glance behaviour data from a wearable, high precision eye-tracking device (Tobii Glasses 3) with subsequent annotation of glance locations on static and dynamic areas of interest (e.g., looking through windscreen / fixating a pedestrian)
- Vehicle speed (recorded from vehicle CAN bus)
- Manual annotation of attributes characterizing the current driving situation (e.g., type of current driving manoeuvre, type of right of way regulation, existence of VRUs close to the ego vehicles driving path)

All data has been sanity-checked, synchronized and downsampled to 10 Hz for the planned data analyses.

3. Results

3.1 Speed-dependency of long offroad glances

Off-road glance durations towards the object "Tablet" have been calculated in accordance with ISO 15007 (International Organization for Standardization, 2020). Data from all reference drivers at speeds above 10 km/h have been combined resulting in n=1055 glances on the tablet device, predominantly from the SuRT condition (see **Figure 2**). The distribution of glance durations varied in the range of 0 to 8 seconds (M= 1.47, Median: 1.24, 95% Percentile: 3.22). Driver-specific Pearson correlations between "tablet glance duration" and "vehicle speed" was found to be negligible or even slightly positive, ranging from r = -0.02 (p=0.71, N= 389) to r = 0.24 (p=0.03, N=85).

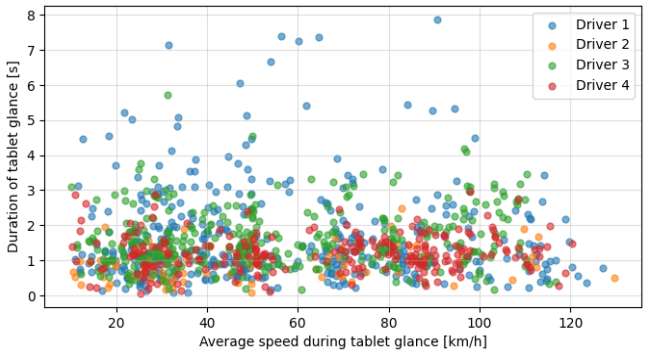


Figure 2: *Scatter plot of individual glance durations on tablet and relationship with vehicle speed*

3.2 Impact of driver's role on glance behaviour

As an exemplary indicator of glance behaviour, proportions of mirror glance time (left, centre, right) relative to the total duration of the drive have been calculated for the conditions "max" and "man" (see **Figure 3**). Paired t-tests for each mirror reveal significantly higher percentages of mirror glance time for left mirror (T=3.90; p=0.03; d=-1.95) and right mirror (T=3.68; p=0.04; d=-1.84) when supervising an L2 automation feature. Due to large interindividual differences in centre mirror usage the descriptive difference for centre mirror usage was not statistically significant (T=2.66; p=0.08; d=-1.33).

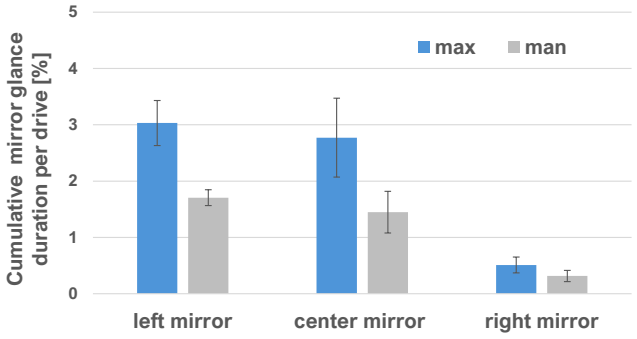


Figure 3: *Comparison of mirror usage in different experimental conditions (mean values with standard errors).*

3.3 Fixation of safety-relevant road scene elements

For this analysis six safety-relevant road scene elements have been selected to further investigate the glance behaviour of reference drivers: oncoming traffic (N=400), pedestrian on sidewalk (N=121), speed limit sign (N=206), stop/yield sign (N=54), green traffic light (N=152), red traffic light (N=105) resulting in a total number of 1038 episodes where these elements were visible in the driving scene. **Figure 4** shows the percentage of "no fixation", "cumulative fixations < 1 second" and "cumulative fixations > 1 second".

Depending on the type of road scene element the percentage of non-fixations ranges from 24% to 93%.

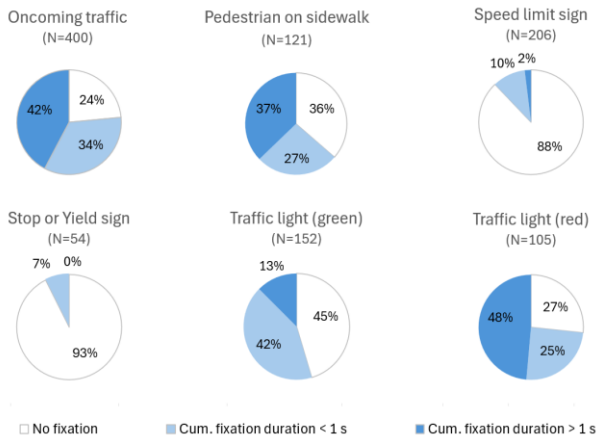


Figure 4: *Cumulative foveal fixation duration on selected road scene elements during visibility*

4. Discussion and conclusion

The glance behaviour of experienced driving instructors, serving as a reference for supervising a non-perfect driving automation system, provides interesting insights:

Although driver distraction monitoring systems typically include vehicle speed as a moderating factor for warning timing, e.g., demanded by regulation (UN ECE R171) or recommended by researchers (see Herbers et al., 2023), the analysis did not reveal a straightforward relationship between vehicle speed and observed eyes-off-road time. Reference drivers seem to have a different / more refined policy to manage secondary task engagement, e.g. by considering the situational demands and risks of the driving situation at hand.

While it is commonly suggested that glance behaviour during L2 supervision should mirror that of manual driving, our data suggest that reference drivers may actually increase their monitoring efforts compared to manual driving as indicated by the increased use of mirrors.

Finally, being attentive and situationally aware does not necessitate foveally fixating all safety-relevant road scene elements. The high percentages of non-fixations are surprising and should be taken seriously when designing future driver monitoring systems that might have the capability to resolve glance directions on an object-level.

The insights gained from this study can serve as valuable input to design advanced algorithms for driver state assessment. Ideally, accurate warning algorithms should be tuned to avoid false alarms (e.g., by not warning reference drivers) while remaining highly sensitive to detect inappropriate glance behaviour.

5. Acknowledgments

This work is a result of the joint research project STADT:up (reference number: 19A22006A). The project is supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), based on a decision of the German Bundestag. The authors are solely responsible for the content of this publication.

6. References

Ahlström, C., Georgoulas, G., & Kircher, K. (2021). Towards a context-dependent multi-buffer driver distraction detection algorithm. *IEEE Transactions on Intelligent Transportation Systems*, 23(5), 4778–4790.

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics*, 68, 138–145. https://doi.org/10.1016/j.apergo.2017.11.010

Boos, A., Feldhutter, A., Schwiebacher, J., & Bengler, K. (2020). Mode Errors and Intentional Violations in Visual Monitoring of Level 2 Driving Automation. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–7. https://doi.org/10.1109/ITSC45102.2020.9294690

Herbers, E., Miller, M., Neurauter, L., Walters, J., & Glaser, D. (2023). Exploratory Development of Algorithms for Determining Driver Attention Status. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/00187208231198932

International Organization for Standardization. (2019). *Road vehicles — Ergonomic aspects of transport information and control systems — Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems* (ISO/TS 14198:2019).

International Organization for Standardization. (2020). *Road vehicles — Measurement and analysis of driver visual behaviour with respect to transport information and control systems* (ISO 15007:2020).

Kircher, K., & Ahlstrom, C. (2017). Minimum required attention: A human-centered approach to driver inattention. *Human Factors*, 59(3), 471–484.

Klauer, S. G., Guo, F., Sudweeks, J., & Dingus, T. A. (2010). *An analysis of driver inattention using a case-crossover approach on 100-car data*. (No. HS-811 334).

Koniakowsky, I. M., Forster, Y., Naujoks, F., Krems, J. F., & Keinath, A. (2023). How Do Automation Modes Influence the Frequency of Advanced Driver Distraction Warnings? A Simulator Study. *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 111–114. https://doi.org/10.1145/3581961.3609888

Marberger, C., Manstetten, D., & Klöppel, C. (2019). Highly automated driving in the real world-A Wizard-of-Oz study on user experience and behavior. *VDI-Berichte*, 2360.

Mueller, A. S., Reagan, I. J., & Cicchino, J. B. (2021). Addressing Driver Disengagement and Proper System Use: Human Factors Recommendations for Level 2 Driving Automation Design. *Journal of Cognitive Engineering and Decision Making*, 15(1), 3–27. https://doi.org/10.1177/1555343420983126

United Nations Economic Commission for Europe (2023 November 8). *Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regard to Driver Control Assistance Systems* https://unece.org/sites/default/files/2023-11/ECE-TRANS-WP.29-GRVA-2024-02e.pdf

Yandel V.(2016, January 22) SuRT - Mobile 1.0. https://apkpure.com/de/surt-mobile/de.lapoehn.surt

Beyond the Source of Distraction: Shared Neural Signatures of Internal and External Distraction in Drivers

Roja Tahamtan [1*], Zahra Razaghi[2], Ali Osia [1]

[1]InnoBrain AB, Department of Mobility, Enhagsvägen 18, 187 40 Täby, Sweden
[2]Center for Nanoscience and Nanotechnology, Institute for Convergence Science and Technology, Sharif University of Technology, Tehran, Iran
*roja.tahamtan@innobraintech.com

Abstract: Driver distraction remains a critical threat to road safety, encompassing both external stimuli and internal mind-wandering. This study investigated neurophysiological correlates of distraction during simulated driving using electroencephalography (EEG). EEG signals were recorded from 21 participants while they performed a continuous driving task, with instructions to detect and respond to police cars that appeared by pressing a button. Three types of audio distractions were presented: internal-focus (prompting self-referential thoughts), non-task-relevant (irrelevant background sounds), and task-relevant (driving-related but secondary information) stimuli.
Reaction times (RTs) to police car detections were used to classify attentional states on a trial-by-trial basis: the 25th percentile (fastest RTs) defined the 'Attentive' state, while trials exceeding the 75th percentile (slowest RTs) defined the 'Distracted' state. Power spectral density (PSD) was analyzed in theta (4-8 Hz) and alpha (8-12 Hz) bands across three regions of interest: temporal, frontocentral, and parieto-occipital.
Paired t-tests with false discovery rate (FDR) correction revealed that the Distracted state exhibited significantly lower theta power in all three regions, alongside higher alpha power compared to the Attentive state. Complementary single-trial repeated-measures correlations (rmcorr) corroborated these findings, demonstrating linear associations between delayed RTs and increased alpha power coupled with decreased theta power across all regions ($p < 0.001$), with the strongest correlations observed in the temporal region.
These results highlight distinct oscillatory signatures of distraction - reduced theta (linked to diminished cognitive engagement) and elevated alpha (indicative of reduced cortical activation) - offering potential biomarkers for real-time monitoring of driver inattention.

1. Introduction

Driver distraction is a major threat to road safety and may arise from both external stimuli and internal mind-wandering (Lee et al., 2008; Forster & Lavie, 2014). In driving, both forms of distraction compete with task-relevant processing and can delay responses to critical events (Taatgen et al., 2021; Voinea et al., 2025). Previous research has shown that mind-wandering is associated with attentional decoupling and measurable changes in neural oscillations, suggesting that EEG may provide a continuous index of moment-to-moment attentional state (Baird et al., 2014; Smallwood & Schooler, 2015).

In this study, we investigated whether alpha- and theta-band EEG activity can differentiate attentive and distracted states during simulated driving. We further examined whether these neural measures vary systematically with trial-by-trial reaction time. A key contribution of this work is showing that distraction across the studied conditions is associated with a common oscillatory pattern characterized by increased alpha and reduced theta activity.

2. Method

EEG data were obtained from the BCIT Mind Wandering dataset (Touryan et al., 2022). After excluding two participants due to excessive artifacts, the final sample comprised 19 participants with normal or corrected-to-normal vision and no reported neurological or psychiatric disorders. Participants performed a continuous simulated driving task in which they maintained lane position and speed while detecting police cars and responding via button press. During driving, they were exposed to three 30-minute auditory conditions: task-relevant, non-task-relevant, and internal-focus audio.

Behavioral performance was assessed using reaction time (RT) and response accuracy to police-car targets. A repeated-measures ANOVA showed no significant differences across auditory conditions for either RT or accuracy. Therefore, trials from all conditions were pooled for EEG analysis.

To maximize contrast between attentional states, trials were classified using RT quartiles: the fastest 25% were labeled Attentive, and the slowest 25% were labeled Distracted. For each target, a 5-second post-stimulus EEG segment was extracted, divided into 1-second epochs, cleaned using an automated artifact rejection pipeline derived from the proprietary signal-processing framework of the InnoBrain Human Cognitive Monitoring Platform (Osia et al., 2025), and averaged to compute PSD. Analyses focused on theta (4–8 Hz) and alpha (8–12 Hz) activity across three regions of interest: frontocentral, temporal, and parieto-occipital. Mean PSD values were compared between attentive and distracted states using paired t-tests with FDR correction. In addition, repeated-measures correlation was used to test trial-by-trial associations between RT and band power within individuals.

3. Results

3.1 *Performance evaluation*

The initial behavioral analysis using a repeated-measures ANOVA revealed no significant main effect of auditory condition on overt performance. Specifically, no significant differences were found across the three auditory scenarios for either reaction time, $F(2, 32) = 0.40$, $p = 0.673$, $\eta^2G = 0.011$, or response accuracy, $F(2, 32) = 2.40$, $p = 0.107$, $\eta^2G = 0.015$ (Fig. 1). Pairwise comparisons likewise showed statistically non-significant (ns) differences across conditions, supporting the pooling of trials for subsequent neural analyses. Fig. 1 shows mean reaction time and response accuracy across the three auditory conditions.

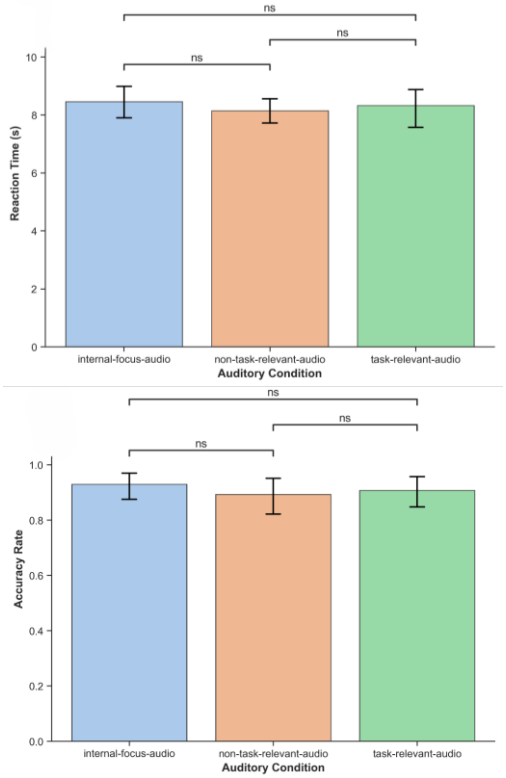


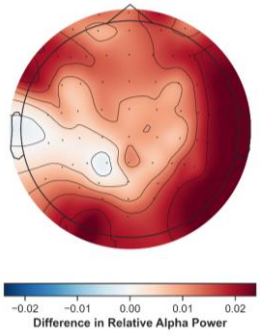
Fig. 1. *Behavioral performance across auditory conditions, (a) Reaction time, (b) Response accuracy.*

3.2 *PSD distribution*

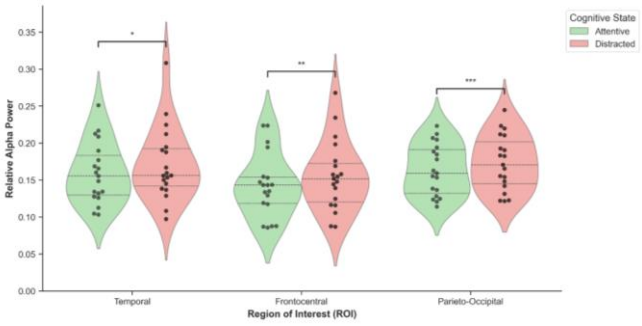
Paired-samples t-tests were conducted to compare relative Alpha and Theta power between the Attentive and Distracted states, with all p-values corrected using the FDR procedure.

Relative Alpha power was significantly higher in the Distracted state across all analyzed regions: parieto-occipital, $t(18) = 4.19$, $p_{fdr} = 0.001$, $d = 0.27$; frontocentral, $t(18) = 3.42$, $p_{fdr} = 0.004$, $d = 0.25$; and temporal, $t(18) = 2.78$, $p_{fdr} = 0.012$, $d = 0.25$ (Fig. 2).

In contrast, relative Theta power was significantly lower in the Distracted state across all regions: temporal, $t(18) = -5.22$, $p_{fdr} < 0.001$, $d = 0.37$; frontocentral, $t(18) = -4.59$, $p_{fdr} < 0.001$, $d = 0.31$; and parieto-occipital, $t(18) = -3.59$, $p_{fdr} = 0.003$, $d = 0.36$ (Fig. 3).

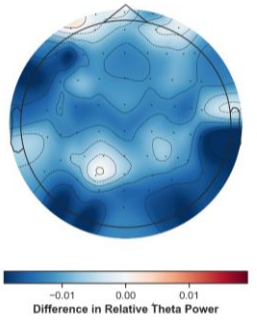


(a)

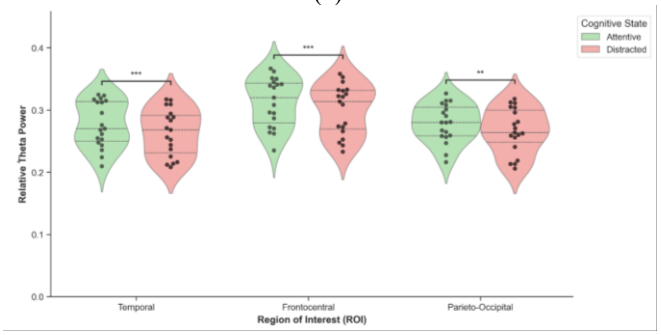


(b)

Fig. 2. *Alpha band (8–12 Hz) oscillatory dynamics differentiating Attentive and Distracted cognitive states.* ***(a)*** *Scalp topography of relative Alpha power (Distracted − Attentive). Red spectrum indicate increased Alpha synchronization during the distracted state,* ***(b)*** *Comparison of relative Alpha power across Temporal, Frontocentral, and Parieto-Occipital regions of interest (ROIs). Internal dashed lines represent the quartiles and median, while black dots represent mean values per participant. Asterisks denote statistical significance following FDR correction (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).*



(a)



(b)

Fig. 3. *Theta band (4–8 Hz) oscillatory dynamics differentiating Attentive and Distracted cognitive states.*

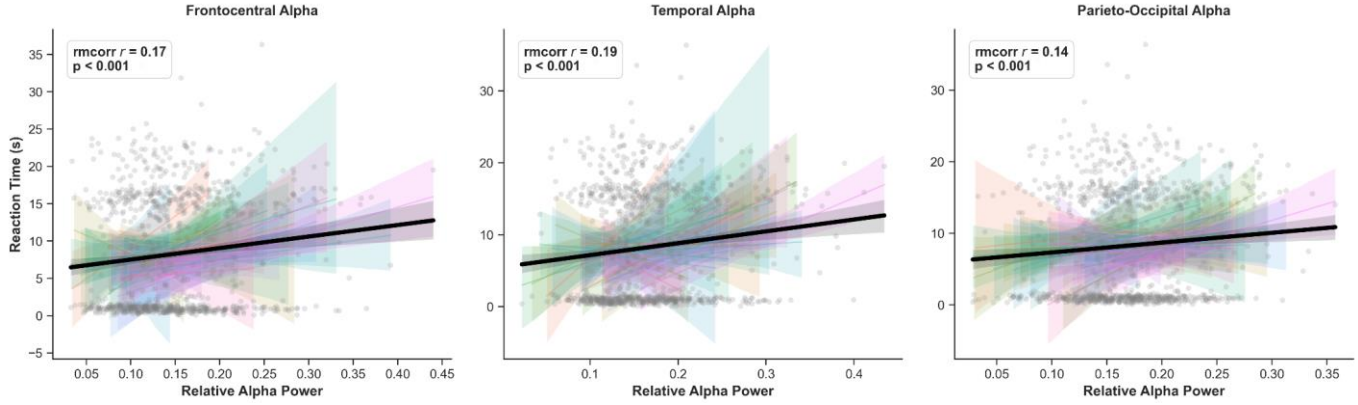
(a) *Scalp topography of relative Theta power (Distracted − Attentive). Blue spectrum indicate a widespread suppression of Theta activity during the distracted state,* ***(b)*** *Comparison of relative Theta power across Temporal, Frontocentral, and Parieto-Occipital ROIs.*
*Asterisks denote statistical significance following FDR correction (** p < 0.01, *** p < 0.001).*

3.3 Trial-by-Trial Brain-Behavior Relationship

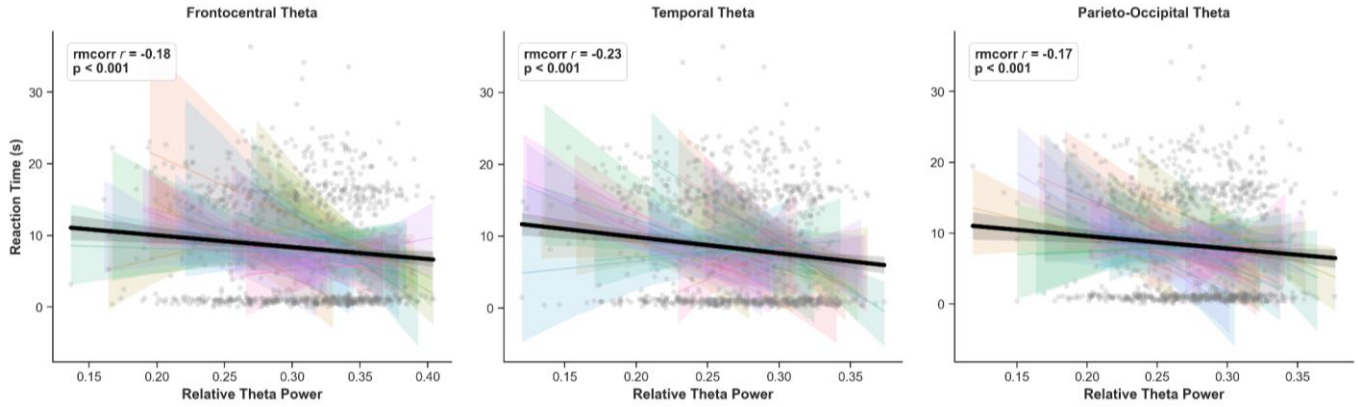
To assess the single-trial association between neural activity and behavioral performance while accounting for intra-individual variance, repeated-measures correlation (rmcorr) was performed between reaction time and relative Alpha and Theta power.

Relative **Alpha** power showed a significant positive correlation with reaction time across all regions, indicating that higher Alpha power was associated with slower responses. The strongest correlation was observed in the Temporal region, *r*rm(933) = 0.188, $p < 0.001$, followed by the Frontocentral, *r*rm(941) = 0.171, $p < 0.001$, and Parieto-Occipital regions, *r*rm(941) = 0.144, $p < 0.001$ (Fig. 4a).

Relative **Theta** power showed a significant negative correlation with reaction time across all regions, indicating that lower Theta power was associated with slower responses. Again, the strongest correlation was found in the Temporal region, *r*rm(933) = -0.232, $p < 0.001$, followed by the Frontocentral, *r*rm(941) = -0.184, $p < 0.001$, and Parieto-Occipital regions, *r*rm(941) = -0.168, $p < 0.001$ (Fig. 4b).



(a)



(b)

Fig. 4. *Trial-by-trial brain-behavior relationships using repeated-measures correlation. Scatter plots depict the single-trial relationship between reaction time (s) and relative power in the* ***(a)*** *alpha frequency bands and,* ***(b)*** *theta frequency bands, across Frontocentral, Temporal, and Parieto-Occipital ROIs.*
Translucent-colored-lines and shaded confidence intervals represent individual participant data fits, while the solid-black-line represents the common within-individual regression slope. The analysis reveals highly significant positive linear associations between reaction time and Alpha power, alongside negative associations with Theta power, across all examined regions.

4. Discussion

These findings are consistent with earlier work linking mind-wandering and distraction to reduced engagement with task-relevant stimuli and altered neural processing (Baird et al., 2014; Smallwood & Schooler, 2015). They also extend prior driving research showing that distraction undermines safe performance and hazard response (Lee et al., 2008; Voinea et al., 2025). The observed increase in alpha and decrease in theta suggest a shift toward reduced cortical engagement and weaker top-down attentional control during distracted driving. Importantly, because behavioral performance did not differ across auditory conditions, the results suggest that distraction across the studied conditions may converge on a common neural pattern, rather than reflecting entirely separate attentional states. This is a key contribution of the study.

5. Conclusion

In conclusion, this research highlights that driver inattention is not merely a transient behavioral lapse but a distinct neurophysiological state characterized by an alpha-theta power shift. These findings have critical implications for the future of human-factors engineering and automotive safety. The identified neural markers provide a highly reliable, unconstrained foundation for the development of passive, real-time Brain-Computer Interfaces (BCIs) and advanced driver-assistance systems (ADAS) capable of continuously monitoring driver cognitive states and mitigating the risks associated with distracted driving.

6. Acknowledgments

This work was funded by InnoBrain AB and conducted between the InnoBrain Neuro-AI team and academic researchers. The EEG signal- preprocessing pipeline used here is based on InnoBrain's proprietary Human Cognitive Monitoring Platform.

References

Baird, B., Smallwood, J., Lutz, A., & Schooler, J. W. (2014). The decoupled mind: Mind-wandering disrupts cortical phase-locking to perceptual events. *Journal of Cognitive Neuroscience, 26*(11), 2596–2607.

Forster, S., & Lavie, N. (2014). Distracted by your mind? Individual differences in distractibility predict mind wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(1), 251–260.

Lee, J. D., Regan, M. A., & Young, K. L. (2008). Defining driver distraction. In M. A. Regan, J. D. Lee, & K. L. Young (Eds.), *Driver distraction: Theory, effects, and mitigation* (pp. 31–40). CRC Press.

Osia, A., Tahamtan, Z., Zhao, L., Davari, M., & Nybacka, M. (2025). A real-time unconstrained EEG-classifier for mental workload monitoring. In *Proceedings of the Driving Simulation Conference (DSC) 2025 Europe XR* (pp. 29–36).

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology, 66*, 487–518.

Taatgen, N. A., van Vugt, M. K., Daamen, J., Katidioti, I., Huijser, S., & Borst, J. P. (2021). The resource-availability model of distraction and mind-wandering. *Cognitive Systems Research, 68*, 84–104.

Touryan, J., Apker, G., Lance, B. J., Kerick, S. E., Ries, A. J., McDowell, K., Johnson, T., & Robbins, K. (2022). *BCIT Mind Wandering* (Version 1.0.0) [Data set]. OpenNeuro.

Voinea, G.-D., Gîrbacia, F., Boboc, R. G., & Postelnicu, C.-C. (2025). The mind-wandering phenomenon while driving: A systematic review. *Information, 16*(8), 681.

Calibrating Non-Invasive Impairment Detection Without Intoxication Ground Truth: An Indirect Application Approach

Denny Kopp *, Frank Oberst

Volkswagen AG, Wolfsburg, Germany
*denny.kopp@volkswagen.de

Abstract: Current driver monitoring assessments address non-fatigue impairment and link detection to driver notification and assistance adaptation. We present an end-to-end engineering approach from function concept to application in the target vehicle system and field monitoring and plausibility validation of a multi-category driving under impairment (DUI) function. The function uses time-windowed behavioral features, index-based assessment, and category-specific decision logic to trigger notification and assistance adaptation. A central contribution is an indirect application approach that leverages non-impaired reference drives to define reference ranges, tune decision behavior under explicit acceptance criteria, and operate under limited availability of intoxication-based ground truth.

1. Introduction

Driving under the influence remains a road safety problem. Alcohol strongly increases crash and injury risk, with a non-linear relationship between blood alcohol concentration and crash risk. [1] Illicit drugs and certain prescription medications can also impair driving, and epidemiological studies report elevated crash risk depending on substance class. [2]

Euro NCAP explicitly includes non-fatigue impairment in driver monitoring and links detection to driver warning and assistance adaptation using non-invasive signals. [3] This motivates behavior-based DUI functions that rely on behavioral manifestations rather than direct substance confirmation.

Experimental work reports measurable behavioral differences between sober and substance-affected driving, particularly in gaze and oculomotor measures. Zemblys and Ahlström (2024) report systematic alcohol-related changes on a test track and emphasize that hardware limitations can reduce measurability in target-system camera setups. [4] Göbel et al. (2025) report simulator-based alcohol effects and show benefits of multi-source fusion for distinguishing sober versus alcohol-affected driving in simulation. [5] Hayley et al. (2024) report stimulant-related gaze changes and differences between methamphetamine and alcohol, illustrating that behavioral profiles can vary by substance. [6] Reviews and meta-analyses confirm that risk estimates differ across substance classes and that study heterogeneity and confounding limit direct transfer of findings across setups. Elvik (2013) reports substance-class-specific risk estimates and emphasizes variability and methodological limitations across studies. [2] Kelly et al. (2004) highlight the relevance of multiple-substance use, including alcohol. [7] Buckley et al. (2025) further highlight large variation in design, simulator fidelity, and outcome metrics in alcohol and cannabis simulator studies, which complicates generalization. [8]

Taken together, literature supports the selection of relevant behavioral dimensions, but application and plausibility validation of assessment and decision logic in the target vehicle system remain challenging. Our paper therefore focuses on how to translate a methodological DUI concept into an assessment and decision logic in the target vehicle system.

2. Method

This paper describes an engineering workflow at a methodological level. Figures, categories, and example metrics are schematic and provided for illustration. They do not represent a product specification or the detailed implementation of any production system. The approach provides behavioral indications of impairment rather than confirming substance use. Driver support actions remain assistive and do not replace the driver's responsibility for the driving task.

2.1 Method overview

We implement the DUI function as a processing chain that computes behavioral features in analysis windows, maps them to indices, fuses indices into category levels, and triggers notification and assistance adaptation when decision criteria are met. Figure 1 summarizes three layers.

Recognition layer: Features are computed from available in-vehicle signals per analysis window. They may include gaze-related measures and steering-related measures such as lateral control indicators.

Assessment layer: Features are mapped to indices that quantify deviations relative to a reference range. Because risk and behavioral profiles can vary across substance classes, DUI is treated as a multi-category state with separate category levels for alcohol, drugs, and medications. [2,6,7]

Decision layer: Category levels are compared against decision criteria to trigger driver notification and assistance adaptation consistent with driver monitoring concepts. [3]

The function is designed as a supportive driver monitoring measure. The driver remains fully responsible for vehicle control at all times.

2.2 Challenges for application and validation

The recognition logic is largely literature-informed, but application in the target vehicle system depends on calibrating the assessment and decision logic under field constraints. We summarize six challenges.

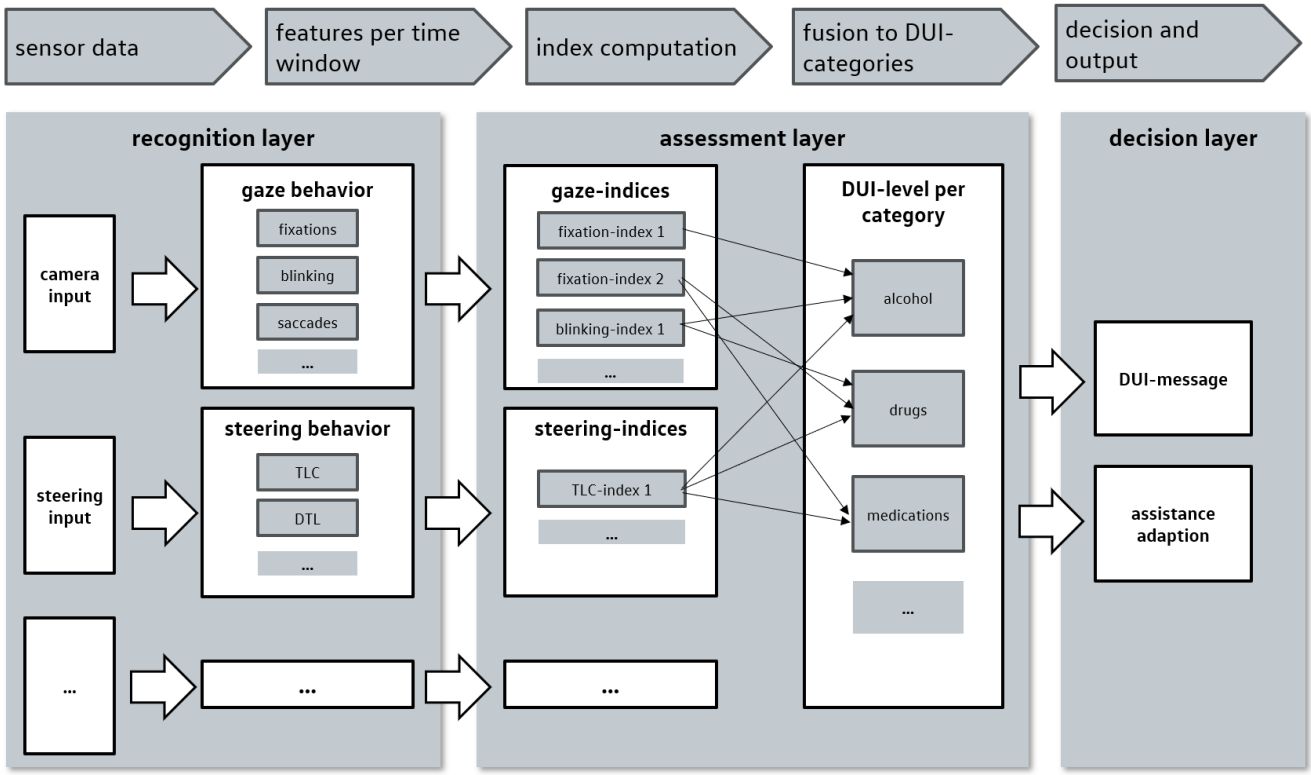


Figure 1 *Overview of the DUI processing chain from windowed feature extraction to index computation, category-level fusion, and decision/output.*

a) Transfer gap
Differences in sensing setups, signal quality, feature definitions, and scenario design limit the direct transfer of published findings into calibration and decision criteria for the target vehicle system. [2,8]

b) System and hardware gap
Sensor configuration, mounting position, data availability, and noise determine which features are measurable and how stable indices are in the target vehicle system. [4]

c) Data scarcity and ground truth
Controlled intoxication studies are practically and ethically limited, especially beyond alcohol, and large ground truth datasets are difficult to obtain. [2,8]

d) Simulator gap
Simulator studies can reveal trends, but variation in fidelity and setups limits their use for calibration in target vehicle system. [8]

e) Coverage and generalization
Robust settings require broad coverage of drivers and scenarios, otherwise decision behavior may not generalize in the field. [8]

f) Category gap
Alcohol evidence is not automatically transferable to other categories, and polydrug contexts can matter, which motivates category-aware evidence and validation strategies. [2,6,7].

2.3 Indirect application approach

We propose an indirect application approach that primarily uses non-impaired reference drives captured with target vehicle system. These drives are available from routine development testing and define the value and variability range of computed features in the target vehicle system. Literature motivates this approach because substance-related impairment can manifest as measurable feature changes compared to sober driving. [4–6] In the indirect approach, deviations from the reference behavior range are treated as behavioural abnormality relevant for impairment indication, while the causal origin remains open.

Controlled studies can support single categories where feasible, and simulator data can be used only if input signals are comparable to target vehicle system. [4,8]

2.4 Feature-wise evaluation and index mapping

For each feature, we define a parameterized acceptance band representing the normal range for assessment. Deviations below or above this band are mapped bidirectionally to two indices that increase gradually with distance from the band boundary (Figure 2). Calibration therefore depends on where the acceptance-band boundaries are placed and when indices start increasing.

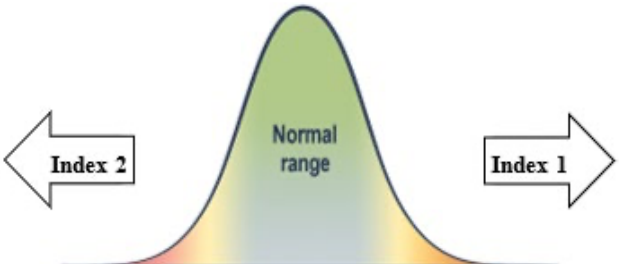


Figure 2 *Schematic example of a bidirectional deviation index for a feature.*

2.5 *Configuration optimization under acceptance criteria*

Because features, indices, and category levels are interdependent, parameterization is performed end-to-end at the overall decision level. In practice, the parameter configuration can be tuned iteratively by evaluating false alerts and other acceptance criteria on the reference dataset and updating parameters until criteria are met or improvements saturate.

The same step can be formulated as constrained optimization to search the parameter space more systematically. Under strong false-alert constraints, solutions can become overly conservative globally or for specific categories. To safeguard triggerability, an optional challenge set can be introduced that contains cases where the system should be able to respond in principle. Challenge cases can be constructed from literature-motivated deviation directions or derived from real driving segments with pronounced deviations. The challenge set does not replace intoxication-based ground truth.

The following variables and parameters are used:

- $\mathcal{D}_0$: non-impaired reference dataset
- $\mathcal{C}$: Challenge-Set
- θ : parameter configuration
- $FP(\theta;\mathcal{D}_0)$: false alert metric
- FP_{min}, FP_{max} : lower and upper bound for $FP(\theta;\mathcal{D}_0)$
- $T^{proxy}(\theta;\mathcal{C})$: triggerability on $\mathcal{C}$
- $g_j(\theta;\mathcal{D}_0)$: acceptance metrics j, with j = 1…J
- g_j^{max}: upper bound for g_j

The problem can be stated as follows:

$$\max \; T^{proxy}(\theta;\mathcal{C}) \tag{1}$$

subject to

$$FP_{min} \le FP(\theta;\mathcal{D}_0) \le FP_{max} \tag{2}$$

$$g_j(\theta;\mathcal{D}_0) \le g_j^{max} \quad \text{for } j = 1\ldots J \tag{3}$$

The objective (1) maximizes T^{proxy}, i.e., triggerability on the challenge set $\mathcal{C}$. The FP constraint (2) bounds the false-alert metric on $\mathcal{D}_0$ to an acceptable range. The additional constraints for g_j (3) represent further acceptance criteria that can be added depending on the application.

2.6 *Field monitoring*

After parameterization we monitor acceptance criteria in field operation. We also run targeted trigger checks to plausibilize that the decision chain can be triggered. Controlled participant studies under influence can complement category-specific evidence where feasible.

3. Results

Table 1 summarizes the analyzed drive dataset from a selected period and its distribution across speed ranges reported as distance and time shares. The dataset includes regular drives and a subset of long drives conducted for fatigue observation. Long drives are included only up to a defined fatigue level for DUI analysis. For feature evaluation, analysis windows are used only if at least 50% of the window provides valid data for feature computation. Otherwise windows are discarded, for example due to low speed, invalid signals, or insufficient driver visibility and tracking quality.

Table 1 *Overview of the analysed drive dataset and its distribution across speed ranges*

	Speed range / drive szenario	*Total*	*Regular drives*	*Long drives*
Number of drives		1513	1407	106
Distance in km	Total	94888	68658	26230
	< 50 km/h	11936	11454	482
	50 - 130 km/h	77636	53043	24593
	> 130 km/h	5317	4162	1155
Travel time in hours	Total	1792	1519	273
	< 50 km/h	880	852	27
	50 - 130 km/h	876	638	237
	> 130 km/h	37	28	9

Figure 3 shows the distribution of an anonymized driver-monitoring metric labeled Feature A over evaluable windows. When all drives are included, an additional accumulation appears in the left value range. When long drives are excluded, this effect is strongly reduced. This indicates that long drives represent a distinct operational profile that can shift feature distributions.

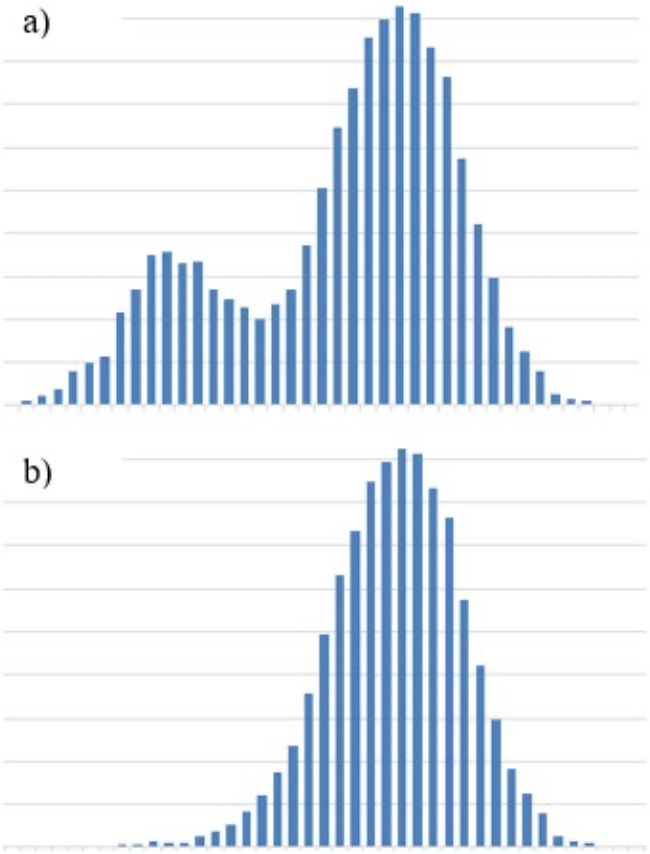


Figure 3 *Histogram of anonymised Feature A over evaluated time windows: (a) all drives, (b) excluding long drives.*

Field monitoring was used to verify that the predefined acceptance criteria were met. Targeted trigger checks at limited scale were conducted to plausibilize the triggerability of the decision chain.

4. Discussion

Our indirect approach enables application of non-invasive DUI and impairment functions in the target vehicle system when intoxication-based ground truth is limited, especially for drugs and medications. [3] Abnormality relative to reference behavior is not substance confirmation. For safety and assistance functions, the exact cause is often secondary to the presence of a driver state that warrants notification or assistance adaptation. This increases the importance of explicit acceptance criteria and robust parameterization to avoid excessive false alerts.

A key limitation is that true positive performance against verified intoxication cannot be fully quantified without controlled ground truth drives. At the same time, the approach reduces dependence on small study cohorts and narrow simulator setups. The long-drive subset demonstrates that naturalistic datasets contain diverse operational profiles that can significantly shift feature distributions, reinforcing the need for broad scenario coverage to improve generalization in the target vehicle system and to reduce overfitting to narrow study setup. Category-based assessment is consistent with substance-class variability and avoids treating alcohol evidence as a proxy for all impairment sources. [2,6,7]

5. Conclusions

We outline a practical end-to-end method to develop and apply the function in the target vehicle system and to support field monitoring under target-system sensing conditions. Literature motivates feature selection, while the main engineering challenge is calibrating assessment and decision behavior under heterogeneous study setups and limited ground truth. [2,8] The proposed indirect application approach leverages naturalistic reference drives to set acceptance bands and tune decision behavior under explicit acceptance criteria, complemented by targeted trigger checks and optional controlled studies where feasible. Results highlight that diverse operational profiles such as long drives can shift feature distributions, which supports broad scenario coverage for generalization.

References

[1] A. K. Høye and I. S. Hesjevoll, “Alcohol and driving—How bad is the combination? A meta-analysis,” Traffic Injury Prevention, vol. 24, no. 5, pp. 373–378, 2023, doi: 10.1080/15389588.2023.2204984.

[2] R. Elvik, “Risk of road accident associated with the use of drugs: a systematic review and meta-analysis of evidence from epidemiological studies,” Accident Analysis & Prevention, vol. 60, pp. 254–267, 2013, doi: 10.1016/j.aap.2012.06.017.

[3] Euro NCAP, “Euro NCAP Protocol – Safe Driving – Driver Engagement v1.0,” Version 1.0, Mar. 2025 (Implementation Jan. 2026). [Online]. Available: https://www.euroncap.com/media/85854/euro-ncap-protocol-safe-driving-driver-engagement-v10.pdf. Accessed: 2026-03-05.

[4] R. Zemblys, C. Ahlström, K. Kircher, and S. Finér, “Practical aspects of measuring camera-based indicators of alcohol intoxication in manual and automated driving,” IET Intelligent Transport Systems, 2024, doi: 10.1049/itr2.12520.

[5] J.-P. Göbel, N. Peuckmann, T. Kundinger, and A. Riener, “Fusion of Driving Behavior and Monitoring System in Scenarios of Driving Under the Influence: An Experimental Approach,” Applied Sciences, vol. 15, no. 10, p. 5302, 2025, doi: 10.3390/app15105302.

[6] A. C. Hayley, B. Shiferaw, B. Aitken, J. Rositano, and L. A. Downey, “Acute methamphetamine and alcohol usage alters gaze behaviour during driving: A randomised, double-blind, placebo-controlled study,” Journal of Psychopharmacology, vol. 38, no. 7, pp. 636–646, 2024, doi: 10.1177/02698811241261024.

[7] E. Kelly, S. Darke, and J. Ross, “A review of drug use and driving: Epidemiology, impairment, risk factors and risk perceptions,” Drug and Alcohol Review, vol. 23, no. 3, pp. 319–344, 2004, doi: 10.1080/09595230412331289482.

[8] L. Buckley, T. Mieran, and K. A. Armstrong, “Rapid review of experimental studies on alcohol and cannabis use and simulated driving behaviour: What can we learn about study methods?,” Transport Reviews, 2025, doi: 10.1080/01441647.2025.2561613.

Driving with future intentions: examining the influence of prospective memory representations on driving performance

Karamjeet Kaur Sarna[12], Azizuddin Khan[2], Sjaan Koppel[1], Amanda Stephens[1], Steven Trawley[3]

[1] Monash University Accident Research Centre (MUARC), Monash University, Melbourne, Australia
[2] Psychophysiology Laboratory, Department of Humanities and Social Sciences, Indian Institute of Technology Bombay, Mumbai, India
[3] Cairnmillar Institute, Melbourne, Australia

*Corresponding author: karamjeet.sarna@monash.edu

Abstract: Driving often requires remembering delayed intentions, such as taking an exit, or locating parking. This ability to remember and execute intended actions in the future is referred to as prospective memory (PM). Laboratory studies show that maintaining such intentions can interfere with ongoing tasks; however, less is known about how these effects manifest during driving. The present study examined how different PM demands influence driving performance by introducing a novel Prospective Memory Parking Paradigm (PMPP). Eleven participants completed four conditions: a baseline driving condition, a focal event-based PM task requiring responses to explicit parking signs, a non-focal event-based PM task requiring drivers to locate parking spaces, and a time-based PM task requiring a button press after two minutes. Driving performance was assessed using vehicle-following measures including modulus, coherence, and lateral lane position. Significant differences across conditions indicated reduced vehicle-following stability when PM demands were present. PM accuracy differed significantly across conditions, with the highest performance in the non-focal condition. However, this improved accuracy was accompanied by reduced vehicle-following coherence and greater lateral lane displacement relative to other conditions. This pattern suggests that the cognitive demands associated with non-focal PM may compromise vehicle control. Time-based PM also demonstrated relatively strong accuracy, reflecting reliance on external clock monitoring. The results align with capacity-based theories of PM, highlighting the trade-offs between executing delayed intentions and driving performance. This research extends laboratory findings to a more ecologically valid context and suggests that delayed intentions may be overlooked cognitive distractors with implications for road safety.

1. Introduction

Remembering to execute planned intentions at the right moment in the future is a vital cognitive skill referred to as prospective memory (PM) (Einstein & McDaniel, 1990; Ellis,1996). A defining characteristic of PM is that delayed intentions are typically embedded within an ongoing activity. Therefore to simulate this in laboratory settings, PM paradigms typically embed a PM task within an ongoing task (Einstein & McDaniel, 1990). This dual-task structure has raised a central question in the PM literature: does retrieving a delayed intention while engaged in an ongoing task impose costs on ongoing task performance?

To understand the processes underlying these costs, The preparatory attentional and memory processes (PAM) theory proposed that the delay between encoding and execution of intention always recruits non-automatic and capacity-consuming processes for PM retrieval (Smith, 2003; Smith & Bayen, 2004). In the absence of strategic monitoring, the intended action is likely to be missed. However, the multiprocess framework accounts for retrieval occurring spontaneously, where the intention is automatically triggered upon encountering a PM cue, or through strategic monitoring (Einstein et al., 2005). Heathcote et al. (2015) proposed an alternative interpretation of the PM costs to reflecting a self-imposed strategic slow down on ongoing task responding to allow PM response to be executed cautiously. This allowed more time for PM-related information to accumulate, thus preventing the ongoing task response from overriding the PM task.

The key tenets shared by all these theories is that PM retrieval requires resource-demanding processes which include monitoring for the target event and engaging preparatory attentional mechanisms, draw from a finite pool of cognitive resources. In this context, subsequent studies demonstrated that the addition of a PM task to an ongoing activity imposed increased cognitive demands on individuals (for e.g., Guynn, 2003; Smith, 2003; Hicks et al., 2005; Marsh et al., 2006; Smith et al., 2007; Strickland et al., 2017).

While PM literature has extensively demonstrated that monitoring for PM cues induces ongoing-task interference in controlled laboratory paradigms such as lexical decision tasks their applicability in more dynamic, high-stakes contexts such as driving—remains less explored. To date, only a limited number of studies have examined PM-related interference in driving contexts. For example, Stephens et al. (2013) observed a descriptive trend suggesting a potential PM-related cost associated with maintaining delayed intentions during vehicle-following, although the effect did not reach statistical significance. In the same vein, Trawley et

al. (2017) provided stronger empirical evidence that PM-related cognitive demands impair driving performance, demonstrating that event-based PM tasks produced greater driving disruption than time-based PM tasks. Building on this, the present study introduces a Prospective Memory Parking Paradigm (PMPP) to investigate how different types of PM representations influence driving behaviour while participants perform a vehicle-following task.

2. Method

2.1 Participants

Participants were recruited from Monash University through flyers placed around the campus. 11 participants were included (mean age = 32.2 years, SD = 8.8), consisting of 6 male and 5 female participants who also received a $ 20 voucher as compensation for their time. The study procedures were approved by the Monash University Human Research Ethics Committee (MUHREC; Project No. 49557).

2.2 Experimental Design

This study employed a within-subjects, repeated-measures experimental design. The independent variable was PM condition, with four levels: no-PM (baseline), focal event-based PM, non-focal event-based PM, and time-based PM. The order of conditions was counterbalanced and no-PM condition was always administered first to avoid carryover effects of PM intentions.

2.3 Apparatus and scenario

Driving tasks were completed using a STISIM® Drive driving simulator (STISIM 100WS). The scenario simulated a 4 km and 14.64 m wide urban route featuring pedestrians, surrounding traffic, oncoming vehicles, road signs, roadside parking, and shops on both sides of the roadway. Throughout the 8 minute drives, the posted speed limit was 40 km/h.

2.4 Prospective Memory Parking Paradigm (PMPP)

This paradigm follows the dual-task framework widely used in prospective memory research. Vehicle-following served as the primary driving task, requiring participants to maintain a safe distance behind a lead vehicle by adjusting their speed accordingly. Embedded within this ongoing driving task were prospective memory (PM) components related to parking. Four conditions were included: (1) Baseline (No PM): participants performed only the vehicle-following task; (2) Focal PM: participants responded to explicit parking signs; (3) Non-focal PM: participants searched for available parking spaces along the road; and (4) Time-based PM: participants executed a response after a specific time interval.

3. Results

3.1 PM Performance

PM performance was calculated as the proportion of correctly executed responses relative to the total number of PM targets presented in each drive. For focal and non-focal conditions, a response was considered correct if it occurred within ±150 m of the target location. For the time-based condition, a hit was recorded when the response occurred within ±30 s of the target time; responses outside these windows were scored as misses.

A repeated-measures ANOVA revealed a significant effect of PM condition on PM accuracy, $F(2, 20) = 16.30$, $p < .001$, $\eta^2_p = .62$. Post hoc comparisons indicated that PM accuracy was significantly higher in the non-focal condition than in the focal condition, $t(10) = 11.74$, $p < .001$, and higher in the time-based condition than in the focal condition, $t(10) = 3.35$, $p = .022$. Overall, PM accuracy was highest in the non-focal condition, followed by the time-based condition, and lowest in the focal condition.

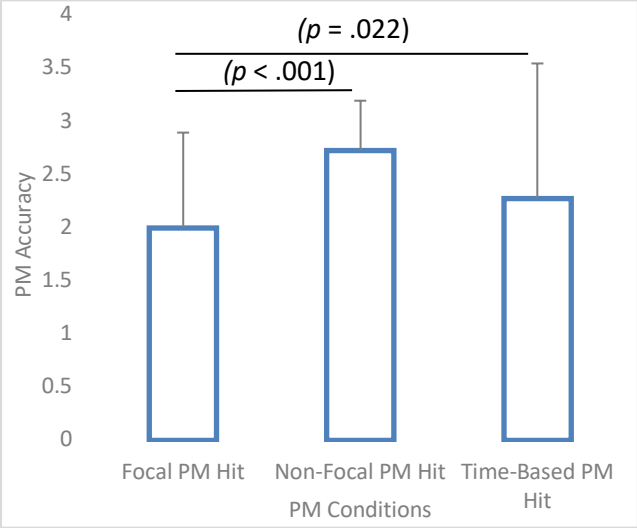


Fig. 1. . Proportion of correctely executed PM responses

3.2 Driving Performance

Vehicle-following measures (coherence, modulus, and time delay) computed by STISIM, along with other driving performance indicators, were analysed using repeated-measures ANOVAs. Greenhouse–Geisser corrections were applied when sphericity assumptions were violated, and Bonferroni-adjusted post hoc tests were conducted ($p < .05$).

Significant effects of PM condition were observed for modulus, $F(2.44, 24.45) = 3.41$, $p = .041$, $\eta p^2 = .254$, and coherence, $F(2.58, 25.76) = 4.96$, $p = .010$, $\eta p^2 = .331$. Post hoc comparisons indicated lower coherence in the non-focal PM condition compared to the no-PM condition ($p = .041$), suggesting less consistent speed adjustments relative to the lead vehicle.

A strong effect was found for mean lateral lane position, $F(2.20, 22.04) = 10.20$, $p < .001$, $\eta p^2 = .505$, with greater lateral displacement in the non-focal PM condition compared to focal ($p = .044$) and time-based conditions (p

< .001). No other driving measures showed significant effects as shown in Table 1.

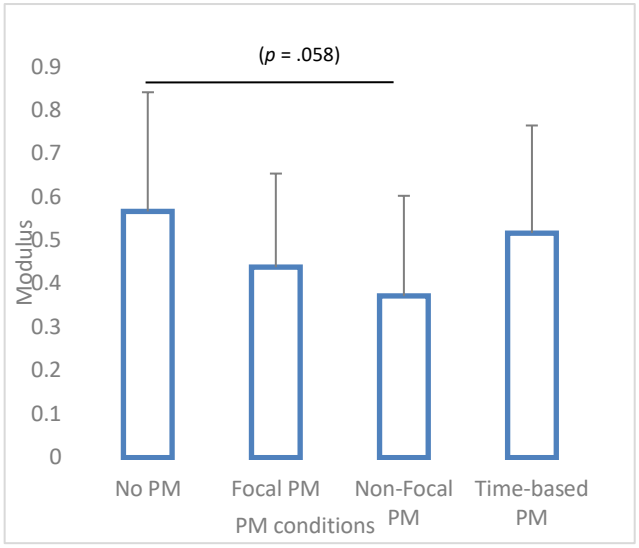


Fig. 2. *Mean modulus across PM conditions.*

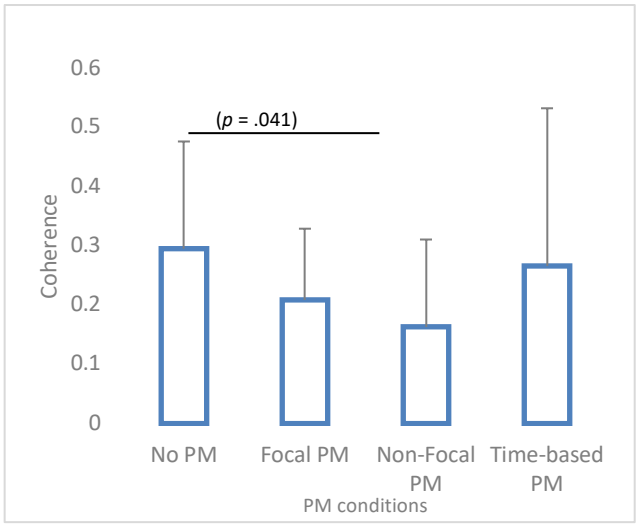


Fig. 3*. Mean coherence across PM conditions.*

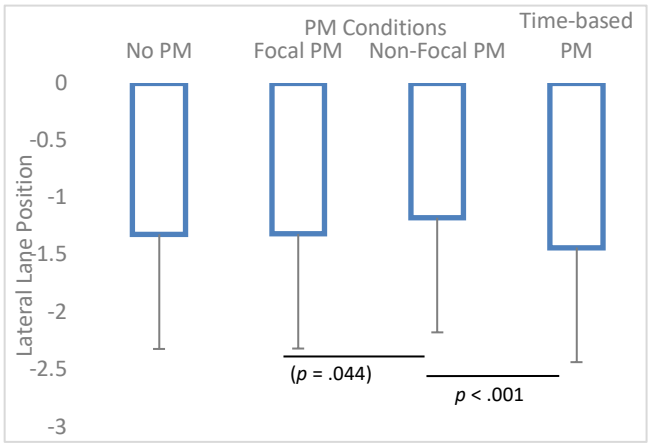


Fig. 4. Fig. 3*. Mean lateral lane position across PM conditions.*

4. Discussion

The present study examined how different types of prospective memory (PM) intentions affect driving performance using a simulated Prospective Memory Parking Paradigm (PMPP). By embedding PM tasks within a vehicle-following task, the paradigm allowed us to assess both the impact of delayed intentions on driving behaviour and the relative cognitive demands of different parking-related intentions.

PM accuracy was highest in the non-focal condition, followed by time-based and focal conditions. This pattern diverges from classic laboratory findings that typically report superior performance for focal PM tasks (Einstein & McDaniel, 2005; Scullin et al., 2010). One possible explanation is that detecting available parking spaces required active environmental scanning, encouraging deliberate and sustained monitoring. In contrast, explicit parking signs may have been detected more passively.

Consistent with our hypothesis, non-focal PM demands were associated with greater impairments in driving performance. Reduced coherence and modulus indicated less stable speed matching with the lead vehicle, and greater lateral lane displacement suggested poorer lane control. These findings support theoretical accounts proposing that non-focal PM tasks require strategic monitoring and therefore impose greater cognitive load. The pattern aligns with prior work showing that event-based PM tasks can disrupt simulated driving (Trawley et al., 2017).

Overall, the study demonstrates that delayed intentions embedded within driving tasks may compete with attentional resources required for vehicle control, highlighting prospective memory as a potentially overlooked source of cognitive distraction in real-world driving.

5. Conclusions

This present study introduced a novel Prospective Memory Parking Paradigm (PMPP) to examine how different types of PM demands influence driving performance in a simulated vehicle-following task. By embedding parking intentions within an ongoing driving task, the study served dual purpose both the cognitive costs of PM on driving and the relative attentional demands of different parking-related intentions. Although PM accuracy was highest in the non-focal condition, this improvement was accompanied by reduced vehicle-following coherence and greater lateral lane deviation, suggesting that strategic monitoring prioritised intention execution at the expense of vehicle control. Consequently, searching for available parking spaces appeared more cognitively demanding than responding to explicit parking signs. Overall, the findings extend laboratory PM interference effects to an ecologically valid driving context, highlighting how monitoring demands may subtly compromise vehicle control.

References

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 16(4), 717–726. https://doi.org/10.1037/0278-7393.16.4.717

Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science, 14*(6), 286–290. https://doi.org/10.1111/j.0963-7214.2005.00382.x

Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., & Breneiser, J. (2005). Multiple processes in prospective memory retrieval: Factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General, 134*(3), 327–342. https://doi.org/10.1037/0096-3445.134.3.327

Ellis, J. A. (1996). Prospective memory or the realization of delayed intentions: A conceptual framework for research. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 1–22). Lawrence Erlbaum Associates.

Guynn, M. J. (2003). A two-process model of strategic monitoring in event-based prospective memory: Activation/retrieval mode and checking. *International Journal of Psychology, 38*(4), 245–256. https://doi.org/10.1080/00207590344000178

Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review, 122*(2), 376–410. https://doi.org/10.1037/a0038952

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language, 53*(3), 430–444. https://doi.org/10.1016/j.jml.2005.04.001

Strickland, L., Heathcote, A., Remington, R. W., Loft, S., & Greene, R. L. (2017). Accumulating evidence about what prospective memory costs actually reveal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(10), 1616–1629. https://doi.org/10.1037/xlm0000400

Trawley, S. L., Stephens, A. N., Rendell, P. G., & Groeger, J. A. (2017). Prospective memory while driving: Comparison of time- and event-based intentions. *Ergonomics, 60*(6), 780–790. https://doi.org/10.1080/00140139.2016.1214288

Marsh, R. L., Hicks, J. L., & Cook, G. I. (2006). Task interference from prospective memories covaries with contextual associations of fulfilling them. *Memory & Cognition, 34*(5), 1037–1045. https://doi.org/10.3758/BF03193250

Scullin, M. K., McDaniel, M. A., & Einstein, G. O. (2010). Control of cost in prospective memory: Evidence for spontaneous retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 190–203. https://doi.org/10.1037/a0017732

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 347–361. https://doi.org/10.1037/0278-7393.29.3.347

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 756–777. https://doi.org/10.1037/0278-7393.30.4.756

Smith, R. E., Hunt, R. R., McVay, J. C., & McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 734–746. https://doi.org/10.1037/0278-7393.33.4.734

Stephens, A. N., Murphy, G., & Trawley, S. L. (2016). Anger and prospective memory while driving: Do future intentions affect current anger? In *Driver Behaviour and Training: Volume VI* (Chapter 7). Ashgate.

Appendix A

Table 1

Means (M), Standard Deviations (SD), and Repeated-Measures ANOVA Results for Driving Measures Across PM Conditions

	No PM		Focal PM		Non-Focal PM		Time-based PM					
Driving measures	M	SD	M	SD	M	SD	M	SD	df	F	p	eta_p2
Modulus	0.56	0.27	0.43	0.21	0.37	0.23	0.51	0.24	(2.44, 24.45)	3.41	0.041	0.254
Time delay	0.96	2.7	-2.12	3.01	-2.15	3.09	-0.44	3.6	(2.66, 26.56)	2.33	0.103	0.189
Coherence	0.29	0.18	0.2	0.12	0.16	0.14	0.26	0.18	(2.58, 25.76)	4.96	0.01	0.331
Speed deviation	2.03	0.49	1.9	0.53	2.03	0.41	2.02	0.4	(1.65, 16.49)	0.387	0.646	0.037
Steering deviation	4.05	4.01	3.45	1.01	6.57	10.05	3.81	3.27	(1.16, 11.62)	0.774	0.416	0.072
Throttle deviation	0.65	0.25	0.71	0.32	0.67	0.19	0.74	0.23	(2.23, 22.33)	0.411	0.69	0.039
Brake deviation	0.85	0.4	0.83	0.44	0.74	0.33	0.9	0.38	(2.03, 20.34)	0.405	0.676	0.039
Lateral lane position	-1.31	0.34	-1.31	0.33	-1.17	0.25	-1.43	0.24	(2.2, 22.04)	10.2	< .001	0.505
standard deviation of lateral position	0.48	0.18	0.49	0.2	0.46	0.13	0.46	0.14	(1.98, 19.77)	0.654	529	0.061

Note. Greenhouse–Geisser corrections were applied where the assumption of sphericity was violated. Reported degrees of freedom are corrected values. ηp^2 = partial eta squared. Significant effects are indicated at $p < .05$. Post hoc comparisons were Bonferroni-adjusted

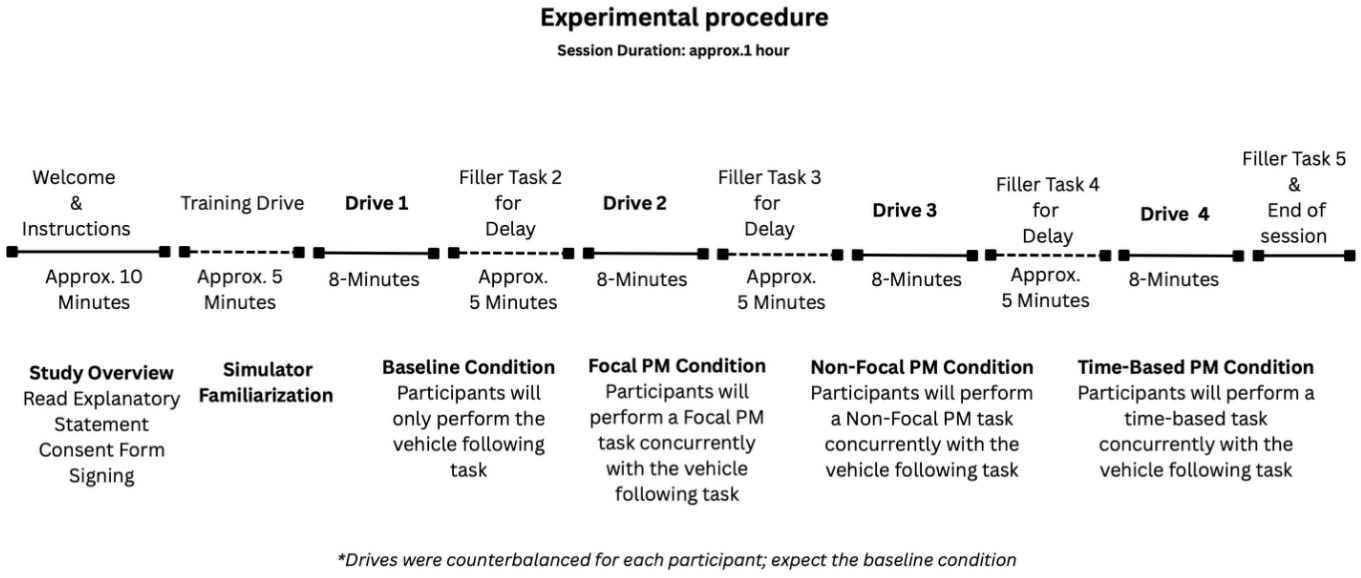


Fig .5. Description of experimental session

Seeing Through Occlusion: Validation of the Occlusion Test Protocol for IVIS Interactions

Rachael A. Wynne[1*], Rita M. Dixon [1]

[1] The University of Newcastle, Discipline of Psychological Sciences, University Drive, Callaghan, Australia
[*]Rachael.Wynne@newcastle.edu.au

Abstract: The National Highway Traffic Administration (NHTSA) Occlusion Test Protocol (OTP) is a globally recognised safety guideline used to measure, understand, and reduce visual distraction caused by in-vehicle technologies. Despite its adoption by vehicle manufacturers, designers, researchers, industry groups and road safety regulatory bodies, there is little empirical evidence confirming the validity of the OTP against real-world glance behaviours. Participants performed a series of tasks within an in-vehicle infotainment system, under three an occlusion method using visual occlusion goggles, participant-controlled occlusion, and no occlusion with eye-tracking. The hypothesis that the OTP would show absolute validity in real-world driver glance behaviour was not supported. Participants glanced off-road less frequently and for shorter durations in surrogate driving tasks than when than is assumed by the OTP. This suggests that the protocol is possibly overestimating the time off-road that could lead to distraction. Additional analyses compare the method of interaction with the system and median time of gaze off-road, again suggesting that the current protocol standard of 8 cumulative 1.5 second glances may not represent the real-world glance behaviours and distraction thresholds. Based on findings of this validation, recommendations for an evaluation of the current OTP are presented which would provide an improvement to the assessment of driver safety

1. Introduction

The NHTSA OTP (NHTSA, 2014) has emerged as a globally accepted tool to understand the impact of visual and cognitive distractions on drivers. This protocol is widely used to evaluate how in-vehicle infotainment system (IVIS) use influences driver attention and safety by using occlusion testing to simulate real-world glances away from the road (Aust et al., 2018; Domeyer et al., 2014; Perez et al., 2013). The occlusion method offers a cost-effective and practical alternative to more complex driving simulations, and provides insights into the visual attention demands that secondary in-vehicle tasks such as using IVIS, places on drivers (Hungund et al., 2021). According to the OTP, an IVIS task should not exceed eight driver eye glances, with each eye glance lasting no more than 1.5 seconds, or 12 seconds of cumulative glance time per task (Pournami et al., 2015). Studies have shown that occlusion testing can provide reliable insights into driver distraction, but limitations such as participant variability and task-specific differences should be considered (Grahn et al., 2023). The OTP is based on research confirming historic beliefs that single long glances away from the road while conducting secondary tasks provide a more consistent estimate of crash risk than total time of eyes off the roadway (Simons-Morton et al., 2014).

The OTP was developed through a collaboration of researchers and engineers, guided by experts in the field of human factors and ergonomics (Hungund et al., 2021; Perez et al., 2013). It specifies that in-vehicle secondary tasks that require a driver's visual attention should be designed to minimise the time the driver's eyes are diverted off the road. The OTP was developed with the goal of reducing driver distractions from in-vehicle tasks, and it ensures that the completion of each task needs only brief and safe visual glances (Perez et al., 2013). However, despite the critical role of NHTSA's OTP in providing safety guidelines for manufacturers and developers of in-vehicle technology, there appears to be a lack of direct experimental evidence supporting the specific time intervals prescribed by the OTP (Grahn et al., 2023; Pournami et al., 2015). Research for this project was also unable to find evidence that the OTP had undergone rigorous scientific validation, and this raises concerns about whether the specified standards are fully supported by empirical evidence. While the OTP provides an essential framework for minimising driver distraction, its current guidelines might not fully reflect the complexities of modern IVIS capabilities. As vehicular technology, specifically IVIS technology advances, the OTP's estimate of duration and frequency of driver off-road glance behaviour while utilising the IVIS may no longer reflect real-life driver interactions with these systems. This gap highlights the need for further research to validate and refine OTP standards to ensure that the protocol remains aligned with the evolving demands of vehicle technology and driver behaviour (Perez et al., 2013).

This study aims to validate NHTSA's OTP, evaluate its current applications, and test its effectiveness as a tool for measuring driver glance behaviour. The extended abstract includes only the results of one analysis, additional analyses to be covered in the full paper.

2. Method

2.1 Participants

Forty licensed regular drivers (25 female, 15 male, M_{age}=35.35 years, *SD* = 12.26), with a minimum five years of solo licensure (*M* = 17.16, *SD* = 10.91) participated in the study. Thirty-two participants reported regular, daily IVIS interaction while driving. The study received ethics approval from the University Human Research Ethics Committee (H-2023-0056).

2.2 Design

The study employed a 3x2 within-subjects factorial design, to determine the impact on driver off-road glance duration. The study manipulated visual occlusion ("glance" representing the OTP standardised occlusion of 1.5 seconds TSOT; participant-controlled occlusion, where participants held the occlusion goggles for as long as they required to do the task; and "no occlusion" monitored via eye tracking). The study also manipulated task interaction mode: visual-manual voice-controlled mode.

Participants completed three of six possible IVIS tasks commonly performed by Australian drivers; change the radio station; initiate a telephone call; change the song via a media playlist; and receive a text message. Where two modes of interaction were possible (visual-manual and voice-control), both modes were performed.

2.3 Materials

This study used Visual Occlusion Goggles (VOGs), eyewear designed to intermittently obscure a user's vision through its LCD lens. Participants press a response button to clear the goggles' visual field temporarily, with the goggles' total shutter open time (TSOT) in this study set to occlude again after 1.5 seconds. When the goggles are transparent, the participant can see the IVIS controls and displays; when occluded, their vision is blocked, preventing them from seeing the task controls and displays. The opaque condition simulates the time a driver is focussing on the primary task of driving.

2.4 ***Procedure***

Participants were instructed to perform a randomised selection of three in-vehicle secondary tasks, which may have included initiating telephone calls, changing media, tuning the radio, and receiving text messages while wearing the VOGs or the eye tracking glasses. While wearing the VOGs, the participants were required to press the response button to clear the goggles visual field temporarily, with the goggles becoming opaque again after 1.5 seconds. The opaque condition intended to simulate the time the participant would be focussing on the primary task of driving.

3. Preliminary Results

Visual occlusion was compared to the OTP standard of driver off-road glance duration of 12 seconds (cumulative) per task, and frequency of eight glances per task. One-sample t-tests revealed significant deviations from the OTP standard across all tasks, ($p < .001$) except for changing a song from a playlist ($p = .419$), which showed moderate evidence for the null ($BF_{10} = 0.32$).

4. Discussion

The aim of the current study was to determine the validity of NHTSA's OTP in assessing real-world driver off-road glance behaviour while conducting secondary in-vehicle tasks, such as those involving IVIS. Contrary to the first hypothesised outcome, results indicated that NHTSA's OTP – defined as eight 1.5 second off-road glances (12 seconds cumulative) per task - does not uphold absolute validity for actual driver glance behaviour. Rather, the actual driver off-road glance durations were found to be shorter than the OTP standard, as evidenced in findings from the "glance" occlusion condition, where five of the six IVIS tasks failed to meet the standard for glance duration, and in the "no occlusion" condition, where no tasks met the duration standard. Similarly, when assessing off-road glance frequency, five out of six tasks failed to meet the standard in both the "glance" and the "no occlusion" conditions.

Contrary to the expected reliability of the OTP regulation standard, findings from the present study suggest that the NHTSA's OTP guideline may be too moderate, and no longer reflective of real-life driver glance behaviour. Glance duration and frequency were both found to consistently fall below the 12 second (cumulative) standard, and findings from the current study suggest that drivers spend less time looking away from the road than was projected when the OTP was established. The results suggest a potential shift in real-life glance behaviour, which may be due to factors that include increased driver awareness of the risks of distracted driving (Oviedo-Trespalacios et al., 2019), and driver self-regulation efforts (Young & Regan, 2013) that can lead to a more cautious attitude when interacting with in-car technologies.

This reduction in driver off-road glance duration to complete secondary tasks supports Young & Regan's (2013) idea that modern drivers cognitively adapt to distraction. Given the cultural shift toward more fast-paced multitasking environments (Strayer et al., 2015), modern drivers may have developed more efficient strategies for managing their interactions with IVIS and other in-vehicle systems. These strategies may result in a reduced need for longer off-road glances due to drivers pre-planning to streamline their interactions and to anticipate future needs and events.

5. Conclusions

This study evaluated the validity of the OTP in reflecting real-world driver glance behaviour while drivers conduct secondary in-vehicle tasks. Contrary to the expected validity of the OTP standards, results showed that driver off-road glances occurred less frequently and for shorter durations than anticipated. This suggests that real-life driving may in fact pose less risk for off-road glances than the OTP anticipates. Nevertheless, as in-vehicle technology available to drivers evolves, it is critical that the OTP regulatory standards and testing methods accurately reflect human distraction tolerances. This emphasises the need for task-specific regulations, and for vehicle manufacturers, policymakers and safety regulators to reconsider how they integrate secondary driving tasks into vehicle systems. Reducing the visual and cognitive demands associated with secondary in-vehicle tasks, such as using IVIS, could result in a dramatic reduction in the risk of serious casualty crashes caused by preventable driver distraction.

References

Aust, M. L., Broström, R., & Lindvall, S. (2018). Assessing in-vehicle secondary tasks with the NHTSA visual-manual guidelines occlusion method. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation* (pp. 373–382). Springer International Publishing. https://doi.org/10.1007/978-3-319-60441-1_37

Domeyer, J., Diptiman, T., & Hamada, H. (2014). Using occlusion to measure the effects of the NHTSA

participant criteria on driver distraction testing. *Transportation Research Part F: Traffic Psychology and Behaviour, 58*(1), 2205–2209. https://doi.org/10.1177/1541931214581463

Grahn, H., Kujala, T., Taipalus, T., Lee, J., & Lee, J. D. (2023). On the relationship between occlusion times and in-car glance durations in simulated driving. *Accident Analysis & Prevention, 182*, 106955. https://doi.org/10.1016/j.aap.2023.106955

Hungund, A. P., Pai, G., & Pradhan, A. K. (2021). Systematic review of research on driver distraction in the context of advanced driver assistance systems. *Transportation Research Record, 2675*(9), 756–765. https://doi.org/10.1177/03611981211004129

NHTSA. (2014). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. In The Federal Register / FIND (Vol. 79, Issue 179, p. 55530) [79 FR 55530]. Federal Information & News Dispatch, LLC. https://www.proquest.com/docview/1562195726/abstract/D4B752B4F2E14AE4PQ/1

Oviedo-Trespalacios, O., Truelove, V., Watson, B., & Hinton, J. A. (2019). The impact of road advertising signs on driver behaviour and implications for road safety: A critical systematic review. *Transportation Research Part A: Policy and Practice, 122*, 85–98. https://doi.org/10.1016/j.tra.2019.01.012

Perez, M. A., Hulse, M. C., & Angell, L. (2013). *Support for NHTSA visual-manual guidelines: Expert review of the visual occlusion method and how it compares to driver eye-glance behavior. Report number DOT HS 811 763* National Highway Traffic Safety Administration. https://www.nhtsa.gov/sites/nhtsa.gov/files/811763.pdf

Pournami, S., Large, D. R., Burnett, G., & Harvey, C. (2015). Comparing the NHTSA and ISO occlusion test protocols: How many participants are sufficient? *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 110–116. https://doi.org/10.1145/2799250.2799255

Simons-Morton, B., Guo, F., Klauer, S., Ehsani, J., & Pradhan, A. (2014). Keep your eyes on the road: Young driver Crash risk increases according to duration of distraction. *Journal of Adolescent Health, 54*(5). https://doi.org/10.1016/j.jadohealth.2013.11.021

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Human Factors, 57* (8), 1300–1324. https://doi.org/10.1177/0018720815575149

Young, K., & Regan, M. (2013). Defining the relationship between behavioural adaptation and driver distraction. In *Behavioural adaptation and road safety: Theory, evidence and action* (pp. 227–242). CRC Press.

Comparing a camera monitor system (CMS) with conventional mirrors in heavy goods vehicles (HGVs)

Victor Fabricius [1*], Nina Theodorsson [2], Daban Rizgary [1], Pontus Larsson [2]

[1] RISE Research Institutes of Sweden, Mobility and Systems, Sven Hultins plats 5, Gothenburg, Sweden
[2] Volvo Group Trucks Technology, Lundby O3, Gothenburg, Sweden
[*]victor.fabricius@ri.se

Abstract: This study compared a camera monitor system (CMS) with conventional mirrors during mixed driving with heavy goods vehicles (HGVs). Ten professional drivers participated in a semi-naturalistic driving experiment providing eye-tracking data, subjective ratings, and general opinions. The central finding is that drivers glanced at the CMS displays more frequently during driving, while subjective workload, eye fatigue, and sleepiness remained similar between mirror conditions. The study provides insights into the current state of CMS and concludes that more nuanced evaluations of driver distraction and inattention may require clearer, context-specific benchmarks for what attention and scanning behavior a given driving situation and task demands.

1. Introduction

This work was conducted within a research project investigating the use and experiences of the existing and future displays, interfaces, and driving environments in cars and trucks. A camera monitor system (CMS) for heavy goods vehicles (HGVs) is increasingly common. However, one of the main motivations for introduction has been reduced drag and fuel consumption and not only an enhanced driving experience. Still, it is expected that a CMS can improve overall driving performance, comfort, and safety (An et al., 2024). Based on this, and drawing on prior interviews with field test drivers, this study was set up as a semi-naturalistic driving experiment comparing CMS and conventional mirrors.

2. Method

Ten test drivers from AB Volvo participated in the study (7 male, 3 female, mean age of 43 years (SD = 16.0, range 25-65)). Their mean driving experience was 15.2 years (SD = 10.2, range 4-36), and they had varying levels of prior CMS experience, ranging from no prior exposure to approximately 4 years.

2.1 Procedure

Starting from Hällered test track outside Borås in Sweden, each participant conducted a driving and reversing task. The driving consisted of a 100 km loop of mixed driving during afternoon (light/dusk conditions, more traffic) and evening (dark conditions, less traffic). The reversing took place inside the test track with the instructions to freely reverse into a parking slot. These tasks were done two times per session, and each participant repeated the same procedure for both mirror conditions (counterbalancing the start with either CMS or conventional mirrors).

2.2 Equipment and data collection

Two Volvo FH trucks with similar drivetrains and trailers attached (pink color for CMS, and orange for conventional mirrors) were prepared with the following equipment: 1) An inverter connected to the onboard battery to power equipment, 2) Three eye-tracking camera sensors with associated computer (Smart Eye Pro, n.d.), 3) Web cameras directed at each mirror, and 4) A laptop with software to collect and synchronize data (iMotions, n.d.). In addition, subjective ratings of mental workload (NASA-TLX), sleepiness (KSS), and visual fatigue (VFS) were collected, as well as general opinions when comparing the two mirror conditions at the end of the session.

3. Results

3.1 Glance patterns

Glances were derived from eye-tracking data using pre-defined areas of interest (AOI) for the left and right CMS and conventional mirrors. During driving, participants spent significantly more time glancing at the CMS displays (see Fig. 1). Since mean glance durations showed similar results between conditions (~450 milliseconds), the differences in dwell time were instead due to more frequent glances at the left display during the CMS condition.

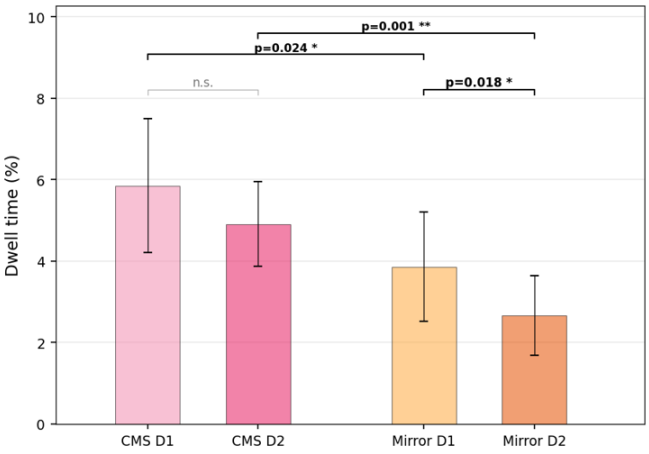


Fig. 1. *AOI dwell time (% of driving task) per condition and drive. Bars = group mean ± 95% CI. Brackets: Holm-corrected post-hoc comparisons (2×2 rm-ANOVA).*

The reverse parking task did not result in any significant differences between CMS and the conventional

mirror conditions, and only a trend of lower dwell time for daytime conditions with conventional mirrors (see Fig. 2). However, this data set was fairly limited and subject to more eye-tracking noise as participants adjusted their seating and head positions more.

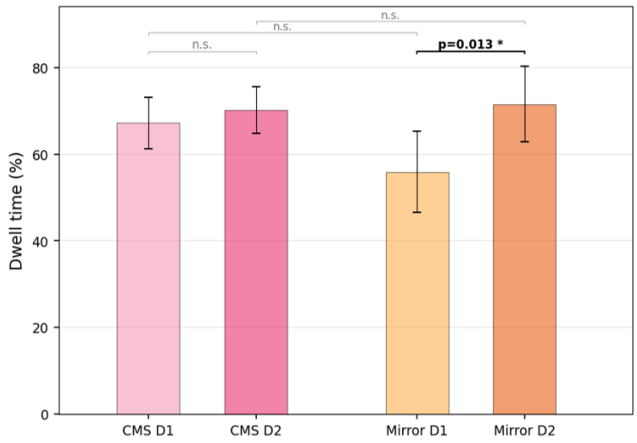


Fig. 2. *AOI dwell time (% of reversing task) per condition and drive. Bars = group mean ± 95% CI. Brackets: Holm-corrected post-hoc comparisons (2×2 rm-ANOVA).*

3.2 Time-on-task

Driving session durations ranged from 96 to 115 minutes across all 40 drives, with a group mean of 104 minutes (SD = 4.1 min). There were also no significant differences in task durations for the reverse parking task (M = 51.8 s, SD = 15.8 s), ruling out that the results on glance behavior would be clearly affected by unequal task durations (see Fig. 3).

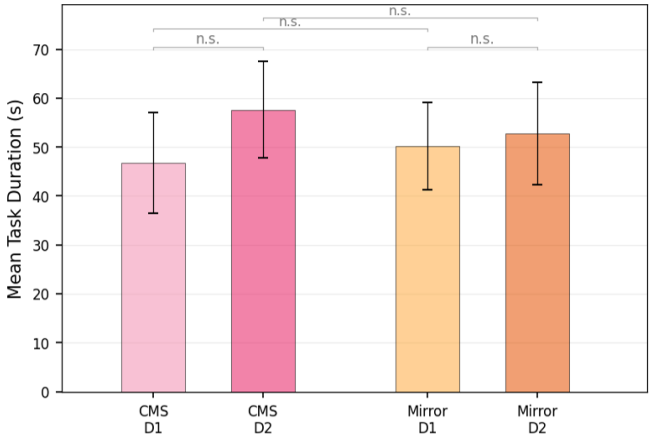


Fig. 3. *Task duration for reverse parking per condition and drive. Bars = group mean ± 95% CI. Brackets: Holm-corrected post-hoc comparisons (2×2 rm-ANOVA).*

3.3 Subjective ratings and opinions

The subjective ratings turned out to be quite consistent between drives and conditions. However, there was a trend of decreased mental demand following the second drive (in darkness but with less traffic). Sleepiness instead increased over the three reporting times during the session (see Fig. 4). While all drivers did not share the same opinions, favored features of the CMS included: 1) Night vision mode, 2) Closer proximity of displays resulting in fewer head turns, 3) View stays cleaner, 4) Better direct field of view (past the A-pillar), 5) Additional functions (e.g., zoom, pan etc.), 6) Lower cabin wind noise, and 7) A narrower cabin when passing wider objects. In terms of possible improvements or weaknesses, these included: 1) Static image perspective when moving your head, 2) More difficult to judge depth and distance, 3) More difficult to reverse (i.e., the trailer looks straight in one screen but not in the other due to a fish-eye lens effect), 4) Lower image quality/resolution and sense of reality, 5) Grainy in darkness (and sometimes difficult to see in rain, fog, and snow), and 6) More difficult to fix if broken.

The quantitative outcome of higher dwell time for CMS during driving, with significantly more glances at the left CMS display, can be linked to comments such as the following: *"The displays are closer to my field of view, so I use them more frequently. They draw focus to themselves." (TP2)*. It was also stated that a lack of experience with CMS could affect their use, with one participant reflecting that *"It is an old habit to use mirrors, and it is easier to reverse straight and judge distances with them"*.

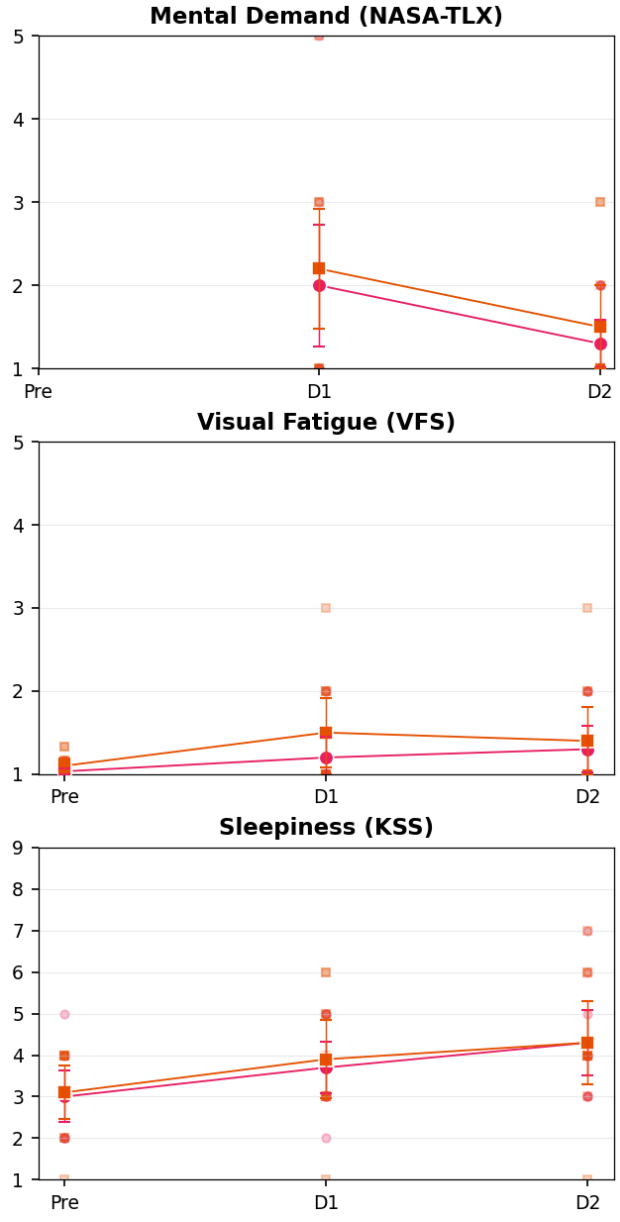


Fig. 4. *Subjective ratings for CMS (pink dots) and mirrors (orange squares). Group mean ± 95% CI connected with trend lines and individual ratings in the background.*

4. Discussion and conclusions

This study compared a CMS with conventional mirrors during mixed driving with HGVs, collecting eye-tracking data, subjective ratings, and driver opinions. The central finding is that drivers glanced at the CMS displays more frequently than at conventional mirrors during driving, while mean glance duration remained similar between conditions. This reflects more frequent sampling of the display rather than longer fixations on it, suggesting drivers are checking the CMS more often rather than being held by it once they look. For the more limited reverse parking sample, this effect was not found. Reversing is a focused, self-paced task whereas driving involves more ambient scanning, and this suggests that the influence of CMS on glance behavior may be task-dependent rather than uniform across all driving activities. These findings speak to a recurring challenge in driver distraction and inattention research of distinguishing genuine distraction from necessary task-relevant information sampling. Unlike an infotainment screen or phone, a CMS display is itself a source of driving-relevant information, so more frequent glances toward it are not distraction in the conventional sense. The question is whether attention drawn to one relevant source comes at the cost of others also required for safe driving. Addressing this will require clearer, context-specific benchmarks for what attention and scanning behavior a given driving situation and task demands.

5. Acknowledgments

The work was partly funded by the Strategic Vehicle Research and Innovation (FFI) Program at Sweden's Innovation Agency (VINNOVA) as part of the SCREENS2 project (grant 2023-00755).

References

An, S., Lee, S., Park, G., Lee, H., Son, M., & Beck, D. (2024). User perception and ergonomic display layout design of truck camera monitor system. International Journal of Industrial Ergonomics, 99, 103547. https://doi.org/10.1016/j.ergon.2024.103547

Smart Eye. (n.d.). Smart Eye Pro - Remote Eye Tracking System. Retrieved March 14, 2026, from https://www.smarteye.se/smart-eye-pro/

iMotions. (n.d.). - Human Behavior Research Platform. Retrieved March 14, 2026, from https://imotions.com/products/imotions-lab/

Comparing Eyes-Off-Road Behavior Between Tactile and Touchscreen Secondary Task Interfaces

John Cliburn [1*], Yi-Ching Lee [1]

[1] George Mason University, Department of Psychology, David King Hall, 4400 University Drive, Fairfax VA, USA
[*] jcliburn@gmu.edu

Abstract: The increasing use of touchscreen interfaces in vehicles has raised concerns about their impact on driver visual attention. This study compared eyes-off-road behavior during secondary task interactions performed using tactile and touchscreen interfaces during a continuous tracking task using a steering wheel and pedals. Eighteen participants completed two counterbalanced tracking sessions while interacting with interface elements in response to randomly presented prompts. Eyes-off-road behavior during each interaction was measured using eye tracking and analyzed with mixed-effects logistic regression. Touchscreen interactions produced dramatically higher odds of off-road glances than tactile interactions ($b = 5.45$, $p < .001$; OR = 232). While 36.1% of tactile interactions were completed without an off-road glance, only 3.5% of touchscreen interactions were completed without an off-road glance. These results indicate that touchscreen-based controls may substantially increase visual distraction during in-vehicle secondary tasks.

1. Introduction

Modern automotive design increasingly favors touchscreen-based controls for in-vehicle information systems (IVIS) over tactile controls (Huo et al., 2024; Suh & Ferris, 2019). While touchscreens provide flexibility and allow many functions to be consolidated into a single interface, they may also increase driver distraction (Crossland et al., 2019; Strayer et al., 2019). Compared with tactile interface elements such as physical buttons and knobs, touchscreen controls have been associated with greater visual, manual, and cognitive workload (Nagy et al., 2023). Driver distraction due to IVIS interactions was estimated to have caused 1.66% of all crashes (Ziakopoulos et al., 2019). Driver distraction more broadly remains a major safety concern, with approximately 13% of police-reported crashes involving distraction (NCSA, 2025) — a figure that could increase as touchscreen interfaces continue to displace tactile controls in modern vehicles.

Total Eyes-off-road time (TEORT) can be used to measure driver distraction due to IVIS interactions because visual attention devoted to IVISs competes with attention needed for monitoring the roadway and responding to hazards (Purucker et al., 2017). Safety guidelines therefore emphasize minimizing off-road glance durations. For example, some guidelines recommend that in-vehicle tasks be designed such that individual glance durations remain below approximately two seconds and that total eyes-off-road time during a task remains limited (NHTSA, 2013).

The present study examined if secondary tasks performed using a touchscreen interface are more likely to require drivers to look away from the roadway than tasks performed using a tactile interface. The study used a within-subjects design in which participants completed repeated secondary task interactions while performing a continuous tracking task in a simulated driving paradigm.

2. Method

2.1 Participants and Design

Eighteen participants were recruited (11 female, 8 male) with a mean age of 25 years old (SD = 5.27). Participants provided informed consent and were compensated with either a $20 gift card or extra credit in an undergraduate course. The experiment used a within-subjects design in which each participant completed two driving sessions: one with the tactile interface and one with the touchscreen interface. Interface order was counterbalanced across participants.

2.2 Apparatus

The tactile interface consisted of a 12” custom control panel including a potentiometer, rotary encoder, two momentary buttons, and two latching buttons connected with a Bodnar BU0836X interface. The touchscreen interface was presented on a 12” touchscreen display and included a slider, two momentary buttons, a scroll bar, and two toggle buttons implemented in Unity. Control types and sizes were matched across modalities and placed in comparable locations.

The primary driving task was developed in Unity and controlled using a Logitech G920 steering wheel and pedals. The task consisted of a vertically scrolling road in which participants tracked a target moving down the center of the road. A Pupil Labs Core eye tracker recording at 60 Hz was used to determine when participants looked away from the primary task display.

2.3 Procedure

Participants practiced the primary driving task for up to five minutes and the secondary task controls on both interfaces for up to 3 minutes. After eye-tracker calibration, participants completed a 15-minute driving session while responding to interface prompts presented at random intervals. Each prompt appeared as a number on the driving display corresponding to a labeled control on the interface (See Fig.

1 for an example). A total of 120 prompts were issued during each session. After a short break to switch interfaces, participants completed a second 15-minute session using the remaining interface.

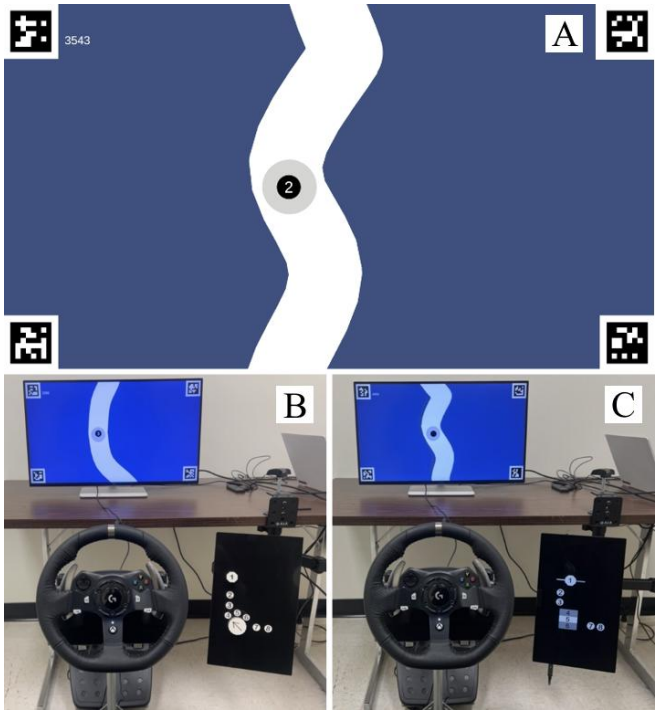


Fig. 1. *(A) Image of the primary task, (B) Tactile secondary task interface, (C) Touchscreen secondary task interface.*

2.4 Analysis

The primary analysis used a mixed-effects logistic regression, as the dependent variable—whether participants looked away from the road during an interaction—was binary and observations were nested within participants.

3. Results

3.1 Interface Effect on Looking Away from Road

A mixed-effects logistic regression model included modality (tactile vs. touchscreen), modality order, and trial number within session (mean-centered) as fixed effects (See Table 1). Random intercepts and random slopes for modality were included by participant to account for individual differences in both baseline tendency to glance off-road and sensitivity to the modality manipulation.

Before interpreting the primary model, a likelihood-ratio test confirmed that allowing modality slopes to vary by participant significantly improved model fit relative to a random-intercept-only model, $\chi^2(2) = 154.28$, $p < .001$, $\Delta AIC = -150.3$, $\Delta BIC = -137.6$. This indicates meaningful between-participant variability in the size of the modality effect, which can also be observed in Figure 2, justifying the more complex random-effects structure. Neither model showed convergence warnings or singularity.

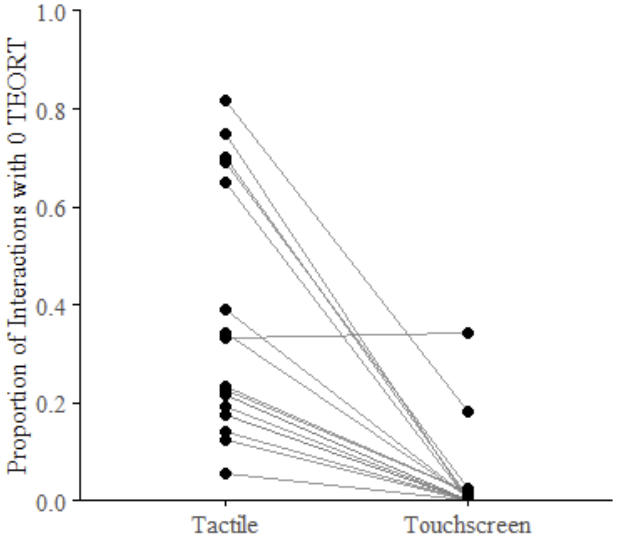


Fig. 2. *Proportions of interactions completed without looking away from the road split by modality for each participant.*

Fixed-effect estimates from the primary model are reported below in Table 1. The random-effects structure yielded participant-level intercept variance of SD = 1.07 and modality slope variance of SD = 2.48 (intercept–slope correlation r = .16).

Touchscreen interactions were associated with substantially greater odds of glancing off-road compared to tactile interactions, b = 5.45, SE = 0.95, z = 5.75, p < .001, OR = 232.00, 95% CI [36.30, 1486.00]. Model-estimated marginal probabilities of glancing off-road were 67.6% [55.7%, 77.6%] for tactile interactions and 99.8% [98.6%, 100.0%] for touchscreen interactions, consistent with the raw descriptive rates of 63.9% and 96.5%, respectively.

Interaction number within session was also a significant predictor of glancing off-road, b = −0.011, SE = 0.001, z = −7.55, p < .001, OR = 0.989, 95% CI [0.986, 0.992]. Each successive interaction was associated with approximately a 1.1% reduction in the odds of looking away, potentially indicating a familiarization effect within the driving sessions. However, modality order was not a statistically significant predictor, b = −0.68, SE = 0.42, z = −1.61, p = .108, OR = 0.51, 95% CI [0.22, 1.16].

Consistent with this difference, Figure 3 shows the distributions of TEORT times split between modalities.

Table 1. *Mixed-effects logistic regression comparing eyes off road occurrence by modality, order, and interaction*

	Estimated Coefficient	Standard Error	Z-value	P-value	Odds Ratio	95% CI
Intercept	1.072	0.334	3.21	0.001	2.92	[1.52, 5.62]
Modality	5.448	0.947	5.75	<0.001	232.39	[36.33, 1486.44]
Order	-0.675	0.420	-1.61	0.108	0.51	[0.22, 1.16]
Interaction #	-0.011	0.001	-7.55	<0.001	0.99	[0.99, 0.99]

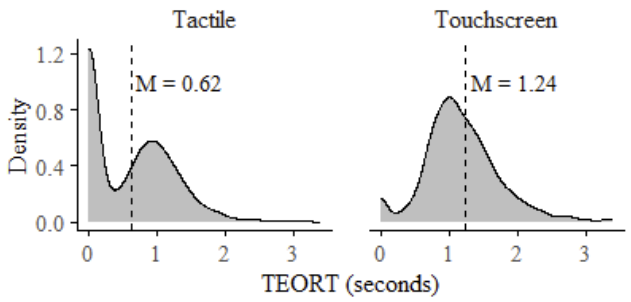


Figure 3. Distribution of Total Eyes-Off-Road Time for Interactions Between Modalities.

4. Discussion

Touchscreen interactions were substantially more likely to require drivers to look away from the roadway than tactile interactions. Nearly all touchscreen interactions required an off-road glance, whereas over a third of tactile interactions were completed without visual diversion. This pattern supports the argument that the absence of tactile feedback increases reliance on visual guidance during interaction. This is consistent with prior work comparing touchscreens and physical buttons, which found that the lack of tactile cues in touchscreen interfaces increases the need for visual confirmation during interaction (Huo et al., 2024).

These findings also have implications for IVIS design guidelines. The high probability of off-road glances observed for touchscreen interactions suggests that certain touchscreen control types could make it more difficult to comply with recommendations that limit glance duration and total eyes-off-road time. Prior research has emphasized the importance of minimizing total eyes-off-road time during in-vehicle interactions because longer visual diversions from the roadway are associated with increased crash risk (Purucker et al., 2017; Strayer et al., 2019).

Several limitations should be considered. The strong modality effect produced extremely large odds ratios, which may be inflated by ceiling effects. In addition, the reduction in eyes-off-road behavior across interactions suggests that continued practice may reduce the need for visual attention over time. However, research indicates that drivers may also exhibit negative behavioral adaptation as they become familiar with in-vehicle technologies. In a multi-day simulator study, drivers' maximum eyes-off-road glance durations increased over repeated exposures to IVIS text entry tasks, suggesting that drivers may become more willing to allocate visual attention to the interface as they grow more comfortable using it (Peng & Boyle, 2015). This pattern suggests that familiarity with an interface does not necessarily reduce distraction risk and may encourage longer visual engagement with in-vehicle systems. Future work could examine how familiarity with an interface influences glance behavior across different control modalities.

5. Conclusions

This study compared tactile and touchscreen interfaces during repeated secondary task interactions in a simulated driving environment. Touchscreen interactions were dramatically more likely to require drivers to look away from the roadway than tactile interactions. These findings highlight the role of tactile feedback in supporting eyes-forward interaction and suggest that touchscreen interfaces may impose greater visual demands during secondary task performance.

6. Acknowledgments

The authors would like to thank the George Mason University Psychology Department for the Summer 2025 Doctoral Research Award that funded the participant recruitment for this experiment.

References

Crossland, A. L., Burnett, G., Large, D. R., & Harvey, C. (2019). The impact of interaction mechanisms with in-vehicle touch screens on task performance. *Contemporary Ergonomics and Human Factors*.

Huo, F., Wang, T., Fang, F., & Sun, C. (2024). The influence of tactile feedback in In-vehicle central control interfaces on driver emotions: A comparative study of touchscreens and physical buttons. *International Journal of Industrial Ergonomics*, *101*, 103586. https://doi.org/10.1016/j.ergon.2024.103586

Nagy, V., Kovács, G., Földesi, P., Kurhan, D., Sysyn, M., Szalai, S., & Fischer, S. (2023). Testing Road Vehicle User Interfaces Concerning the Driver's Cognitive Load. *Infrastructures*, *8*(3), 49. https://doi.org/10.3390/infrastructures8030049

National Center for Statistics and Analysis. (2025). Distracted driving in 2023 (Research Note. Report No. DOT HS 813 703). National Highway Traffic Safety Administration.

National Highway Traffic Safety Administration. (2013). Visual-manual NHTSA driver distraction guidelines for in vehicle electronic devices (Federal Register 78 (81)). In: Author, Washington, DC.

Peng, Y., & Boyle, L. N. (2015). Driver's adaptive glance behavior to in-vehicle information systems. Accident Analysis & Prevention, 85, 93–101. https://doi.org/10.1016/j.aap.2015.08.002

Purucker, C., Naujoks, F., Prill, A., & Neukum, A. (2017). Evaluating distraction of in-vehicle information systems while driving by predicting total eyes-off-road times with keystroke level modeling. Applied Ergonomics, 58, 543–554. https://doi.org/10.1016/j.apergo.2016.04.012

Suh, Y., & Ferris, T. K. (2019). On-Road Evaluation of In-vehicle Interface Characteristics and Their Effects on Performance of Visual Detection on the Road and Manual Entry. Human Factors: The Journal of the Human Factors and Ergonomics Society, 61(1), 105–118. https://doi.org/10.1177/0018720818790841

Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D. J., & Biondi, F. (2019). Assessing the visual and cognitive demands of in-vehicle information systems. *Cognitive Research: Principles and Implications*, *4*(1), 18. https://doi.org/10.1186/s41235-019-0166-3

Ziakopoulos, A., Theofilatos, A., Papadimitriou, E., & Yannis, G. (2019). A meta-analysis of the impacts of operating in-vehicle information systems on road safety. *IATSS Research*, *43*(3), 185–194. https://doi.org/10.1016/j.iatssr.2019.01.003

Prerequisites for Adaptive ADAS: Learning from Fatal Crashes in Sweden

Johan Karlsson[1,2]

[1] Cognitian AB, Vehicle Safety Research, Verdandigatan 5, 59874 Mariannelund, Sweden
[2] SAFER Vehicle and Traffic Research at Chalmers

[*]johan@cognitian.se

Recent developments position driver monitoring systems (DMS) as a mechanism for adapting ADAS intervention thresholds in real time, based on detected driver state. The project examined 224 fatal crashes in Sweden involving driver degradation to assess the theoretical effectiveness of this approach. Crashes were coded according to a new coding scheme, including longitudinal vs lateral origin, error types similar to the familiar omission/commission taxonomy, here expressed as 'underperformance' vs 'overconfidence' and an estimation of whether driver degradation would be possible to detect in time with confidence. Results include descriptive statistics about ten basic scenarios, breakdown of these into longitudinal and lateral origin crashes, how different driver degradations relate to different basic errors, and a breakdown of 'suspected distraction or sleepiness related action' into primarily distraction-related and primarily sleepiness related crashes. It is argued that longitudinal crashes followed by lateral origin crashes, both with non-transient driver degradation patterns are most suitable for adaptive interventions. However, as many of the crashes included vehicles not equipped with AEB or LKA/ELK, the primary recommendation is to first promote full market penetration of standard ADAS. For transient impairments and quickly developing crash scenarios, driver state detection and crash prediction uncertainties compound quickly, leaving little confidence basis for intervention - in the words of Hippocrates: 'first, do no harm'.

1. Introduction

While fatality rates on the roads are getting lower largely thanks to the combination of infrastructure changes such as divided highways and improvements in protective safety, crashes attributed to driver degradation such as sleepiness, distraction, alcohol or drugs are still fairly frequent (Forsman 2015, 2023).

Advanced Driver assistance (ADAS) technology such as Automatic Emergency Braking (AEB) or Lane Keeping Assistance (LKA) has been available for some time, including systems to help maintain lane position, keep distance from vehicles ahead, and apply emergency braking in critical situations (Danielsson 2025). Driver monitoring systems have developed considerably in recent years, with increasingly capable detection of drowsiness, distraction and impairment (Halin et al. 2021). Recent discourse has proposed using these systems to adapt intervention thresholds of existing ADAS. The premise is straightforward: if a driver shows signs of impairment such as sleepiness, distraction, or substance influence, systems could intervene earlier or more forcefully than they would for an alert driver.

This study had two aims. The first was to map scenarios in fatal crashes caused by impaired driving ability due to fatigue, distraction and the influence of alcohol or drugs, in order to better understand how to intervene at an operational level close to the crash. The second aim was to analyse the possible effects of adaptive interventions, such as making automatic braking systems and lane keeping assist systems adaptive to the driver's state.

The premise for this project is that the driver bears primary responsibility for operating the vehicle, whilst collision avoidance systems and driver assistance systems function as complementary protective mechanisms to reduce the risk of or consequences of accidents

In order to better understand the timing and actions involved in crashes with suspected driver impairment, we analysed in-depth data from 224 fatal crashes in Sweden.

2. Method

2.1 In-depth accident meta analysis

The project was granted access to a selection of the in-depth studies of fatal accidents collected by the Swedish Transport Administration during the years 2010-2023, a total of 2,708 accidents with 6,551 involved people, and 2,980 fatalities. From the original database at person level, the following selection criteria were used: traffic participant role: driver. Vehicle year: 2000 or newer was selected to exclude the very oldest vehicles that may not be entirely representative with regard to safety and road characteristics. Accident types - all except wildlife accidents. The weather conditions were no precipitation and road condition was dry or moist roadway to exclude unfavourable road or visibility conditions. Finally, parameters "suspicion of

fatigue-/distraction-related action", "alcohol" or "drugs" should be "Yes". In the end 373 cases were filtered out for analysis, of these a full review of 224 were completed due to time/budget cap. These were distributed according to Table 1. Priority was given to include all cases of accident types with relatively few cases, and then filling up with the two most common types, single-vehicle accidents and head-on collisions. In absolute terms, these accident types are still most common in the study, with 62 and 67 cases respectively.

Table 1 *Number of crashes per type*

Accident type	Reviewed	Number in selection	Percentage covered
Turning	11	11	100
Bicycle/Moped vs vehicle	13	13	100
Pedestrian vs motor vehicle	27	27	100
Train/Tram	5	5	100
Crossing paths	11	11	100
Head-on	67	113	59
Overtaking	7	7	100
Single-vehicle	62	154	40
Rear-end	20	20	100
Miscellaneous	12	12	100
Total	224	373	60

The study utilised post-crash investigation data, including police reports, medical records, toxicology results, and crash reconstruction analyses. Crashes were classified according to two primary patterns of driver degradation: underperformance — late or absent manoeuvring due to missed hazard detection, itself stemming from distraction, sleepiness, or in some cases visual obstruction — and overconfidence — excessive speed for conditions, resulting in loss of vehicle control.

2.2 Analysis framework

Each case was coded by the author against a structured scheme covering error type (omission or commission), impairment temporality (transient or continuous), relevant ADAS type (lateral, longitudinal, or both), and an assessment of whether timely detection of impaired driver capability was considered feasible.

A fundamental limitation is acknowledged: exact event sequences cannot be reconstructed with certainty from post-crash evidence. Classification of degraded driver performance often relied on circumstantial evidence, which tends to confirm stereotypical patterns — the drowsy driver drifting off the road, for instance. Medical evidence and toxicology provide stronger grounds for classification in substance-related cases, but behavioural inference remained necessary throughout. This evidentiary uncertainty is not merely a methodological inconvenience; it mirrors the detection uncertainty faced by real-time monitoring systems attempting to classify driver state from indirect signals.

3. Results

3.1 ADAS type

Most crashes were in the categories head-on and single vehicle. The impairments seen most frequently in those scenarios are alcohol and sleepiness. Head-on crashes had a higher share of suspected sleepiness while single vehicle crashes had a higher share of alcohol and drugs.

3.2 Subdivision of impairments

A subsample of 172 cases classified as 'suspicion of distraction or sleepiness related action' were analysed to be categorised further as primarily related to either factor. Of those cases 45% were interpreted to be related to distraction, 36% to sleepiness, 23% were unclear, and the 6% remaining cases were related to substances or illness.

3.3 Intervention feasibility

A qualitative analysis was performed for each basic scenario, based on most frequent impairment type, impairment being transient or non-transient, scenario origin being predominantly longitudinal or lateral errors, covering the following categories: Scenario Description, Impaired Driver Capability, Scenario Variation and Characteristics, Complexity in Responsibility Distribution, ADAS Potential, Adaptation Possibilities, and Adaptation Challenges. Standing out in the analysis were single-vehicle and rear-end crashes, which were considered to show the greatest potential benefit from adaptive intervention, followed by head-on collisions.

4. Discussion

4.1 Crash origins

It could be argued that most head-on and single vehicle crashes have origin in unrecoverable lane deviation/departure. This could be related to sleepiness being related to ceased manoeuvring while within speed limits while drug and alcohol-related crashes can possibly be related to overconfidence and over-speeding, resulting in loss of control and high speed road departure.

The investigated crashes can be categorised as originating in excessive variation or deviations in either longitudinal (over-speeding, late braking, early starting) or lateral (lane departure leading to head-on collision or road departure and collision with terrain) directions. This generally maps to two or three ADAS systems. For longitudinal direction there is the continuous support from ACC, and the emergency intervention in AEB, for lateral direction there is the

continuous support from LKA and a more intermittent version in LDW.

4.2 Performance degradations

The human driver performance degradations covered can be categorised as transient in nature, such as a brief visual distraction, or non-transient, such as fatigue, sleepiness or substance use. Ironically, some transient states are fairly uncomplicated to detect, while the non-transient impairments show more subtle signs, large intra-individual variations and a small signal-to-noise ratio and are therefore not always straightforward to detect with confidence.

4.3 Predominant errors

The main driver errors encountered in the study were driver underperformance — related to missed detections, or early or late actions, mainly errors of omission, (Reason 1990) and related to distraction, sleepiness, and in some cases alcohol use. These can be contrasted with overconfidence errors which are mainly errors of commission — wrong selection of action or selecting actions outside the current suitable performance envelope based on visibility, road conditions. The overconfidence errors seem to be mostly related to drug use, and in some cases also to alcohol. The error type at least in part determines what kind of intervention is needed.

4.4 Intervention adaptions

The two main intervention adaptation variants considered in this project were 1. earlier longitudinal action, to the extent that it constitutes similarly early intervention as ACC deceleration and 2. stronger LKA torque, in order to keep vehicles in lane until an emergency stop could be performed.

Given the detection uncertainty for both driver impairment and conflict/collision detection, it would make most sense to intervene in the case of a continuous impairment detected with good confidence, for a longitudinal scenario with good time margins, such as a rear-end scenario, but also to some extent adapt LKA torque so as to avoid lane departures.

As a further step one could also consider late adaptation of AEB to a transient impairment such as visual inattention/distraction, but due to the short timespans involved extraordinary caution should be taken not to trigger a hard AEB braking based on insufficient evidence and thereby worsen a situation, essentially the Hippocratic ‘first do no harm’.

5. Conclusions

The DMS-to-ADAS integration pathway deserves careful scrutiny of its benefit-risk profile — including failure modes, false positive rates, and driver adaptation. Many of the crashes analysed were preventable in principle, but many vehicles were not equipped with intervening ADAS, which makes the obvious first step to ensure full market penetration of basic ADAS functions, such as ACC/AEB and LKA.

For prevention of substance-related crashes, interventions that prevent starting the driving task (like alcohol interlocks) show greater promise than graduated real-time threshold adaptation. The latter faces fundamental challenges: limited detection confidence and narrow intervention windows. Proceed with caution.

6. Acknowledgements

This study was conducted within the ADAPTIN project, funded by Trafikverket Skyltfonden (grant 2024/106736).

References

Danielsson, J. (2025). *Fordonsflottans utveckling – Avancerade förarstödsystem och digitala tjänster*. Swedish Transport Administration.

Forsman, Å. (2015). *Alkohol, droger och läkemedel hos omkomna personbilsförare. VTI notat* 11-2015

Forsman, Åsa. (2023). *Förekomst av alkohol och andra droger hos omkomna motorfordonsförare. VTI Rapport* 1185

Halin, A., Verly, J. G., & Van Droogenbroeck, M. (2021). Survey and Synthesis of State of the Art in Driver Monitoring. *Sensors*, 21(16), 5558. https://doi.org/10.3390/s21165558

Reason, J. T. (1990). *Human Error*. Cambridge Univ. Press.

Conversational cognitive load meets lateral hazard exposure: How environment shapes crash risk

Minxiang Zhao[1], Jonas Bärgman [1*], Carol Flannagan [2]

[1] Chalmers University of Technology, Mechanical Engineering, Hörsalsvägen 7A, Gothenburg, Sweden
[2] University of Michigan Transportation Research Institute (UMTRI), Ann Arbor, U.S.
[*]jonas.bargman@chalmers.se

Abstract: This study examines how talking/singing while driving affects crash risk under naturalistic conditions, comparing high- and low lateral-interaction environments. Using weighted SHRP2 data, we evaluated talking-related tasks against four baseline types, including a new Acceptable-task baseline representing societally accepted behaviors. Talking generally did not increase crash risk under realistic baselines, possibly reflecting increased forward glances and preserved automatized braking responses (cognitive control hypothesis), although talking/singing (audience unknown) was elevated in low lateral-interaction environments. In high lateral-interaction environments, modest risk increases appeared, especially for hand-held phone conversations. Visual-manual phone tasks consistently showed elevated crash risk across all contexts. The work underscores the importance of baseline choice and driving environment.

1. Introduction

Many countries have banned phone use while driving due to strong evidence that it harms safety. Numerous studies show that *Visual-manual phone tasks* greatly increase crash risk. Although these *Visual-manual tasks* produce clearly elevated on-road risk in naturalistic driving studies (NDSs), talking on the phone (hand-held or hands-free) has not shown a significant crash-risk increase in real-traffic data (e.g., Klauer et al., 2014). Depending on the *baseline*, several analyses even suggest a protective effect (e.g., Young, 2019). Explanations include: (a) talking may replace riskier tasks, and (b) talking increases forward-directed glances. Critics note that the NDS crashes used for the risk assessments are largely rear-end events in relatively simple situations, where increased eyes-on-road time can be protective. Other critics cite simulator studies showing slower reactions under cognitive load (e.g., Drews et al., 2008). However, yet others argue these simulator effects may reflect study design rather than real-world crash risk (Engström et al., 2017).

We examine how talking or singing affects crash risk under *Low* versus *High lateral-interaction*. Because conversational cognitive load shifts gaze forward, we expect higher risk when hazards arise laterally rather than straight ahead.

We also assess how baseline choice influences crash-risk estimates; prior work shows large differences, with *Ideal driving* yielding the highest—and arguably unrealistic—risk values. As an alternative, we introduce an *Acceptable-task* baseline, consisting of tasks that are societally accepted while driving and unlikely to be legally restricted.

2. Method

2.1 Dataset

We used data from the SHRP2 NDS, in which ~3,400 participants drove instrumented vehicles during everyday use from 2010–2013. The data-acquisition systems recorded video, and on-board sensors, totaling ~56 million km (Hankey et al., 2016). SHRP2 identified and severity-classified safety-critical events. We used crash Levels I–III (excluding low-risk tire strikes) and *baseline* segments, which serve as non-critical comparison samples. Both *crashes* and *baselines* are provided as 6-s segments (epochs). All crash types (rear-end, run-off-road, pedestrian, etc.) were included.

SHRP2 also provides manually annotated secondary-task data (Hankey et al., 2016). These tasks are non-driving activities (such as phone use and eating/drinking) that may divert attention from driving. SHRP2 defines 62 secondary-task types, and each epoch may include 0–3 tasks.

2.2 Covariate and population balancing

We used propensity scores to balance observed driving-context covariates between events with and without the task. Propensity scores—the probability of treatment given covariates—were estimated and used as weights so that events without the task represented the counterfactual for those with it. Because the SHRP2 age distribution differs from that of U.S. licensed drivers, we applied an additional age-based weight using the ratio of U.S. to SHRP2 age-group proportions. The final weight used was the product of the propensity and age components and was used in all analyses.

2.3 Tasks and baseline

2.3.1 Talking/singing tasks

This paper focuses on talking/singing tasks, which are among the most common secondary tasks. SHRP2 defines four talking/singing-related tasks: two passenger-interaction tasks (front-seat and rear-seat interaction) and two non-passenger tasks (*Talking/singing, audience unknown* and *Cell-phone talking/listening, hand-held*). We concentrate on the latter two, as passenger interactions often involve off-road glances toward the passenger, and we want to focus on the cognitive component of the conversational tasks.

2.3.2 Other tasks

In addition to talking tasks, we also evaluate the risks associated with *Radio tuning* and a broader set of primarily *Visual-manual cell-phone tasks* (Table 1). *Radio tuning* is part of the foundation for NHTSA's distraction guidelines and corresponds to what we classify as *Acceptable tasks*. In contrast, *Visual-manual cell-phone tasks* represent actions that are illegal in many jurisdictions.

Table 1 *Visual-manual cell-phone-related tasks (SHRP2 annotation categories)*

- Holding	- Texting
- Browsing	- Dialing hand-held
- Locating/reaching /answering	- Dialing hand-held using quick keys
- Other	

2.3.3 Baselines

To assess crash risk, the tasks need to be compared to an appropriate baseline. We define four baselines:

- *Ideal-driving* ("Ideal") – only tasks directly related to safe driving.
- *Acceptable-driving ("Acc.drv.")* – Ideal-driving behaviors plus societally accepted tasks unlikely to be restricted (see Table 2).
- *Acceptable-task* ("Acc.task") – only the acceptable tasks, excluding driving-related tasks.
- *Task-absent* – includes all tasks except the task being evaluated.

Table 2 *Acceptable-tasks (SHRP2 annotation categories)*

- Adjusting/monitoring climate control	- Drinking from open container
- Drinking with lid and straw	- Drinking with lid, no straw
- Adjusting/monitoring radio	- Drinking with straw, no lid
- Smoking cigar/cigarette	- Biting nails/cuticles
- Passenger in adjacent seat interaction	- Passenger in rear seat interaction

2.3.4 Locality

In addition to glance behavior, SHRP2 records the locality of each event (see Table 3). To assess how context influences crash risk, we grouped localities into *High lateral-interaction* areas—where drivers are more likely to encounter crossing traffic and vulnerable road users and must monitor the surroundings more actively—and *Low lateral-interaction areas*, where driving is more forward-focused (e.g., highways, open country).

Table 3 *Context categories (SHRP2 annotation categories)*

High lateral-interaction	**Low lateral-interaction**
- Moderate Residential	- Bypass/Divided Highway with traffic signals
- School	- Interstate/bypass/ divided highway without traffic signals
- Church	- Open Country
- Business/Industrial	
- Playground	
-Urban	

2.4 Analysis

To evaluate the effect of secondary tasks on crash risk, we use the weighted SHRP2 crash and baseline data to estimate odds ratios (ORs) and their corresponding CIs using logistic regression models.

3. Results

First, the OR comparing *Low* versus *High lateral-interaction* shows that crash odds in *High lateral-interaction areas* were more than three times higher (age-weighted OR = 3.45, 95% CI = 2.79–4.31).

Across all tasks and conditions, ORs follow a consistent pattern (Table 4): risk is highest when using the *Ideal-driving* baseline and decreases as broader baselines are applied, reaching their lowest values with the *Task-absent* baseline.

For *Talking-related* tasks and for *Adjusting/monitoring the radio*, the *Ideal-driving* and *Acceptable-driving* baselines yield significantly elevated ORs. In contrast, when compared with the *Acceptable-task* or *Task-absent* baselines, these tasks generally do not show significantly increased crash risk.

Two notable exceptions emerge. First, for *Hand-held phone talking* in *Low lateral-interaction areas*, the OR relative to the *Task-absent* baseline is significantly below 1 (two crashes only), with the OR relative to the *Acceptable-task* baseline also below 1, though not statistically significant. Second, *Talking/singing, audience unknown* in *Low lateral-interaction areas* shows consistently elevated ORs across all baselines, unlike the pattern for *Hand-held phone talking*.

When the two talking tasks are combined, ORs for the broader baselines (e.g., *Task-absent*) cluster near 1. This reflects the opposing OR patterns of the two subtypes and the higher prevalence of *Talking/singing, audience unknown*. In *Low lateral-interaction areas*, where *Talking/singing, audience unknown* has 14 of the 16 crashes, combined ORs follow it (1.22–1.82).

Across all baselines, combined talking tasks show lower ORs than *Adjusting/monitoring the radio*. By contrast, *Visual-manual cell-phone tasks* (Table 1) exhibit consistently elevated crash risks, with ORs above 2 for all baselines except in *Low lateral-interaction areas*, where confidence intervals are wide.

Table 4 Odds ratios with confidence intervals across tasks, driving context and baseline

Task	Locality	Baseline	OR (95% CI)
Cell Phone Hand Held	All	Ideal	**2.03 (1.33 – 2.98)**
		Acc. drv.	**1.91 (1.26 – 2.79)**
		Acc. task	1.39 (0.89 – 2.10)
		Task abs.	1.23 (0.82 – 1.77)
	High lateral interact	Ideal	**2.31 (1.49 – 3.43)**
		Acc. drv.	**2.19 (1.43 – 3.24)**
		Acc. task	**1.68 (1.06 – 2.60)**
		Task abs.	1.49 (0.98 – 2.17)
	Low lateral interact	Ideal	0.34 (0.01 – 1.85)
		Acc. drv.	0.31 (0.01 – 1.66)
		Acc. task	0.19 (0.01 – 1.09)
		Task abs.	**0.17 (0.01 – 0.93)**
Talking/ singing Audience Unknown	All	Ideal	**1.45 (1.05 – 1.96)**
		Acc. drv.	**1.40 (1.02 – 1.87)**
		Acc. task	1.04 (0.74 – 1.46)
		Task abs.	0.94 (0.70 – 1.25)
	High lateral interact	Ideal	1.21 (0.83 – 1.72)
		Acc. drv.	1.16 (0.81 – 1.63)
		Acc. task	0.86 (0.58 – 1.27)
		Task abs.	0.79 (0.55 – 1.09)
	Low lateral interact	Ideal	**2.50 (1.29 – 4.55)**
		Acc. drv.	**2.47 (1.30 – 4.39)**
		Acc. task	**2.24 (1.04 – 4.84)**
		Task abs.	**1.88 (1.01 – 3.23)**
Cell Phone Hand Held & Talking Audience Unknown	All	Ideal	**1.62 (1.24 – 2.09)**
		Acc. drv.	**1.55 (1.19 – 1.98)**
		Acc. task	1.13 (0.84 – 1.52)
		Task abs.	1.03 (0.80 – 1.30)
	High lateral interact	Ideal	**1.53 (1.14 – 2.03)**
		Acc. drv.	**1.46 (1.10 – 1.92)**
		Acc. task	1.07 (0.77 – 1.49)
		Task abs.	0.99 (0.76 – 1.28)
	Low lateral interact	Ideal	1.82 (0.96 – 3.28)
		Acc. drv.	1.77 (0.95 – 3.10)
		Acc. task	1.42 (0.68 – 2.92)
		Task abs.	1.22 (0.67 – 2.07)
Adjusting/m onitoring radio	All	Ideal	**1.76 (1.02 – 2.83)**
		Acc. drv.	**1.73 (1.01 – 2.77)**
		Acc. task	1.62 (0.91 – 2.74)
		Task abs.	1.16 (0.68 – 1.84)
	High lateral interact	Ideal	1.65 (0.90 – 2.77)
		Acc. drv.	1.64 (0.90 – 2.74)
		Acc. task	1.60 (0.85 – 2.85)
		Task abs.	1.13 (0.62 – 1.87)
	Low lateral interact	Ideal	2.40 (0.54 – 7.08)
		Acc. drv.	2.27 (0.52 – 6.54)
		Acc. task	1.79 (0.38 – 6.19)
		Task abs.	1.50 (0.35 – 4.16)
Cell Phone Tasks (Except Talking/liste ning, hand-held & Dialing hands-free)	All	Ideal	**3.38 (2.57 – 4.41)**
		Acc. drv.	**3.16 (2.42 – 4.07)**
		Acc. task	**2.41 (1.76 – 3.29)**
		Task abs.	**2.38 (1.84 – 3.03)**
	High lateral interact	Ideal	**3.51 (2.61 – 4.67)**
		Acc. drv.	**3.27 (2.46 – 4.30)**
		Acc. task	**2.50 (1.78 – 3.52)**
		Task abs.	**2.50 (1.91 – 3.24)**
	Low lateral interact	Ideal	1.93 (0.75 – 4.22)
		Acc. drv.	1.85 (0.72 – 3.95)
		Acc. task	1.53 (0.55 – 3.77)
		Task abs.	1.38 (0.55 – 2.88)

4. Discussion

In *Low lateral-interaction areas* (e.g., highways), talking on a *Hand-held phone* showed little or no added crash risk, and in some comparisons even appeared protective relative to societally accepted tasks including *Adjusting the radio*. This aligns with earlier NDS work and the cognitive-control hypothesis (Engström et al., 2017), which holds that conversational engagement boosts forward glances without hindering automatized braking to looming hazards.

In *High lateral-interaction areas* (e.g., urban settings with pedestrians and crossing traffic), *Hand-held phone talking* showed moderately elevated risk relative to the *Acceptable-task* baseline. This may indicate that although talking increases forward gaze, it may reduce awareness of lateral hazards that are more common in these areas.

Visual-manual cell-phone tasks showed high crash risk, highlighting that off-road glances and manual interaction remain the most hazardous distractions.

The elevated ORs for *Talking/singing, audience unknown* in *Low lateral-interaction areas* do not match the pattern for *Hand-held phone conversations*. Possible explanations include heterogeneity within the "audience unknown" category or behavioral differences (e.g., simple car-following scenarios versus more complex maneuvers such as lane changes, where cognitive load could have greater consequences). Further analysis of behavioral and glance patterns would be needed to clarify this discrepancy.

Methodologically, the finding that the *Ideal-driving* baseline overstates risk is consistent with prior work: this baseline represents unrealistically task-free driving. The *Task-absent* baseline, by contrast, reflects the activities that talking typically replaces and therefore provides the most policy-relevant comparison. Under this baseline, *Talking tasks* generally yield ORs near 1, while *Visual-manual tasks* remain clearly elevated. The *Acceptable-driving* and *Acceptable-task* baselines are more realistic than *Ideal driving* but are best treated as sensitivity analyses because they do not capture task replacement effects.

Finally, because SHRP2 contains mostly lower-severity crashes, these findings cannot speak to rare high-severity outcomes. Future work using large-scale continuous in-vehicle monitoring may help address this gap.

5. Conclusions

This study indicates that talking or singing while driving generally does not raise crash risk when compared with realistic baselines, particularly for Hand-held phone talking in Low *lateral-interaction* areas where increased eyes-on-road time and preserved automatized responses may mitigate risk. In high-interaction environments, however, talking—especially *Hand-held phone talking*—shows moderately elevated risk. Some inconsistencies warrant caution, but overall the results highlight the importance of baseline choice and driving context when assessing distraction-related crash risk.

6. Acknowledgments

This work was partially funded by the QUADRIS project (grant number 2025-00834) project funded by FFI, but primarily by Chalmers University of Technology.

This research used data from the SHRP 2 Naturalistic Driving Study, hosted by the Virginia Tech Transportation

Institute under contract to the Transportation Research Board. The data are accessed under license SHRP2-DUL-A-3-19-432. The findings and conclusions of this paper are those of the authors and do not necessarily represent the views of VTTI, TRB, or the National Academies.

References

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, 14(4), 392–400. https://doi.org/10.1037/a0013119

Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. *Human Factors*, *59*(5), 734-764. https://doi.org/10.1177/0018720817690639

Hankey, J. M., Perez, M. A., & McClafferty, J. A. (2016). *Description of the SHRP2 naturalistic database and the crash, near-crash, and baseline data sets - task report*. https://vtechworks.lib.vt.edu/bitstream/handle/10919/70850/SHRP_2_CrashNearCrashBaselineReport_4-25-16.pdf;sequence=1

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine*, 370(1), 54-59. https://doi.org/10.1056/NEJMsa1204142

Young, R. A. (2019). *Cell phone conversation and relative crash risk update*. In Encyclopedia of Information Science and Technology (4th ed.). IGI Global. https://doi.org/10.4018/978-1-5225-2255-3.ch521

How Drivers Allocate Visual Attention During Extended Conversational AI Interactions

Chris Monk [1*], Allegra Ayala [1], Christine S.P. Yu [1], Ellie Robbins [1], Gregory M. Fitch [2], Dara Gruber [2]

[1] Exponent, Inc., 1150 Connecticut Ave. NW, Suite 1100, Washington, DC, USA
[2] Google, Inc., 1600 Ampitheatre Parkway, Mountain View, CA, USA
[*] cmonk@exponent.com

Abstract: Modern conversational AI systems enable extended, multi-turn dialogue that may alter how drivers allocate their visual attention, even when aggregate visual demand remains low. This study examined visual attention during hands-free phone conversation and multi-turn conversational AI interaction using a glance-episode-based framework. Eye-tracking data collected during real-world driving were analyzed to characterize the duration, sequencing, and run-level structure of on-road and on-screen glance episodes. On-road glance episodes were substantially longer than on-screen episodes across both tasks, but mean episode duration, episode sequencing, and on-road run duration did not differ meaningfully between interaction types. Visual attention patterns remained stable, with brief device-directed glances embedded within sustained periods of roadway monitoring. These findings suggest that extended conversational AI interaction can be supported by visual attention structures comparable to those observed during conventional hands-free phone use and highlight the value of episode-based analyses for evaluating emerging voice-based in-vehicle technologies.

1. Introduction

Visual distraction is a well-established contributor to crash risk, with naturalistic driving studies showing that prolonged or poorly timed glances away from the forward roadway substantially increase the likelihood of safety-critical events (Klauer et al., 2006; Klauer et al., 2014; Victor et al., 2015). As a result, visual attention metrics play a central role in the evaluation of in-vehicle systems intended to support secondary task engagement during driving.

Design guidance from the National Highway Traffic Safety Administration (NHTSA) and the Alliance of Automotive Innovation emphasizes limits on individual off-road glance duration and cumulative visual demand, leading to widespread use of metrics such as mean glance duration and total eyes-off-road time (Alliance of Automobile Manufacturers, 2006; NHTSA, 2013). While these measures summarize overall visual demand, they do not characterize how visual attention is distributed over time. Tasks with similar cumulative visual demand may differ in safety relevance depending on whether off-road glances are brief and intermittently sampled or occur in clustered sequences that interrupt roadway monitoring (Victor et al., 2015).

Voice-based interfaces generally reduce visual demand relative to visual–manual interaction, but modern conversational systems based on large language models enable extended, multi-turn dialogue that may alter the temporal organization of visual attention even when aggregate visual demand remains low (Reimer et al., 2021; Monk et al., 2023). Understanding these effects requires analyses that move beyond aggregate metrics to examine the structure and sequencing of glance behavior.

The present study focuses on the temporal organization of visual attention during hands-free phone conversation and multi-turn conversational AI (Gemini Live) interaction. Using glance episode and run-based analyses, this work examines how drivers interleave brief device-directed glances with sustained periods of on-road viewing. A glance episode is defined as a contiguous period of visual attention directed either to the roadway or to the in-vehicle display, and on-road runs are defined as uninterrupted sequences of on-road glance episodes bounded by device-directed glances. These constructs allow visual attention to be examined in terms of both individual episodes and sustained periods of roadway monitoring.

The data analyzed here were collected as part of a broader on-road evaluation of conversational AI (Monk et al., 2026), but are re-examined using an episode-based framework that was not emphasized in the prior work.

2. Method

2.1 Participants

Twenty-four licensed drivers (balanced by sex and age across four groups: 18–24, 25–39, 40–54, 55–60) completed the study. Participants reported regular driving experience and routine use of voice-based conversational agents. Eight additional recruits were excluded due to early termination, driving performance issues, or equipment failure. The study was approved by Exponent's Institutional Review Board.

2.2 Apparatus and Measures

Testing was conducted on public roads in Bowie, Maryland using a 2025 Chevrolet Trailblazer driven along a predefined low-complexity route. Visual turn-by-turn guidance was provided via Android Auto using Google Maps. A Google Pixel 9 smartphone running a pre-release version of Gemini Live supported all voice-based interactions through the in-vehicle display.

Eye movements were recorded using Tobii Pro Glasses 3. Scene video and gaze data (see Fig. 1) were processed in Tobii Pro Lab to derive glance episodes to on-road and in-vehicle regions of interest. Cognitive workload was additionally assessed using a tactile Detection

Response Task, and subjective workload and perceived distraction ratings were collected following each task.



Fig. 1. *Forward view from the Tobii Pro 3 Glasses camera..*

2.3 Tasks and Procedure

Five task conditions were evaluated while driving: visual turn-by-turn guidance only (baseline), hands-free phone call, Gemini Live single-turn interaction, Gemini Live multi-turn interaction, and an auditory OSPAN task. During the baseline condition, participants followed visual turn-by-turn guidance only. The phone call involved a brief hands-free conversation with an experimenter outside of the vehicle. Gemini Live single-turn interactions consisted of one prompt–response exchange, whereas multi-turn interactions involved extended dialogue. The OSPAN task served as a high cognitive-load comparison. Participants completed all tasks while driving the same route, with task order counterbalanced.

3. Results

3.1 Mean Glance Episode Durations

Mean glance episode durations were first examined to compare overall visual attention behavior between the Hands-Free Phone Call and Gemini Live mulit-turn conditions. A linear mixed-effects model with task (Hands-Free Phone Call, Gemini Live) and glance episode location (on-screen vs. on-road) as fixed effects and participant as a random effect revealed a pronounced effect of glance episode location. Across both task conditions, on-road glance episodes were substantially longer than on-screen episodes (Fig. 2).

In contrast, no meaningful differences were observed between the two conditions in mean on-road or on-screen glance episode duration. Estimated marginal means for both glance types were comparable across tasks, indicating that average episode duration alone did not distinguish between the two voice-based interaction tasks.

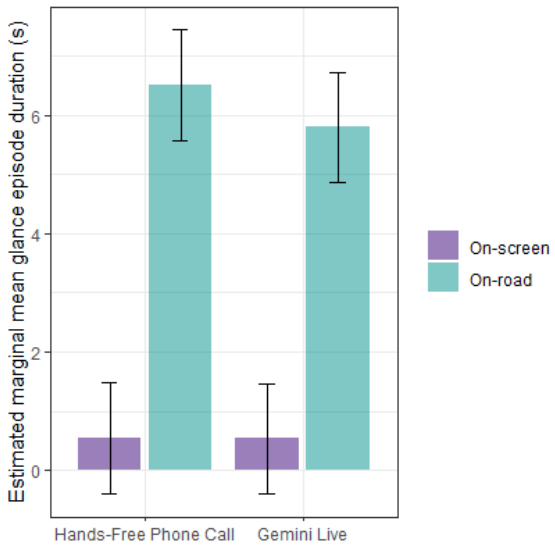


Fig. 2. *Estimated marginal mean glance episode durations by task and glance episode location.*

3.2 Sequential Structure of Glance Episodes

To examine the temporal organization of glance behavior beyond mean differences, glance episodes were analyzed as an ordered sequence within each trial minute. Analyses were restricted to the first three minutes of each trial and to the first 24 glance episodes within each minute, ensuring adequate participant representation at each episode index.

Across both task conditions, on-screen glance episodes remained brief and relatively stable across the episode sequence, whereas on-road episodes were consistently longer (Fig. 3). This pattern was evident across the full range of analyzed episode indices and did not differ systematically between the Hands-Free Phone Call and Gemini Live conditions. Visual inspection of episode-indexed means revealed a consistent alternation between short device-directed glances and substantially longer on-road episodes, with no evidence that on-road viewing became shorter or more interrupted over time.

3.3 On-Road Run Structure Between Device Glances

Because neither mean episode duration nor episode ordering differentiated tasks, a deeper analysis focused on the structure of continuous on-road monitoring between device-directed glances. On-road “runs” were defined as contiguous sequences of on-road glance episodes bounded by on-screen glances, representing uninterrupted periods of roadway monitoring.

A linear mixed-effects model predicting log-transformed on-road run duration with task as a fixed effect and participant as a random effect showed no reliable difference between the task conditions. Back-transformed estimated marginal means indicated similar geometric mean run durations across tasks (Fig. 4), suggesting that drivers maintained comparable periods of uninterrupted roadway monitoring during both types of voice-based interaction.

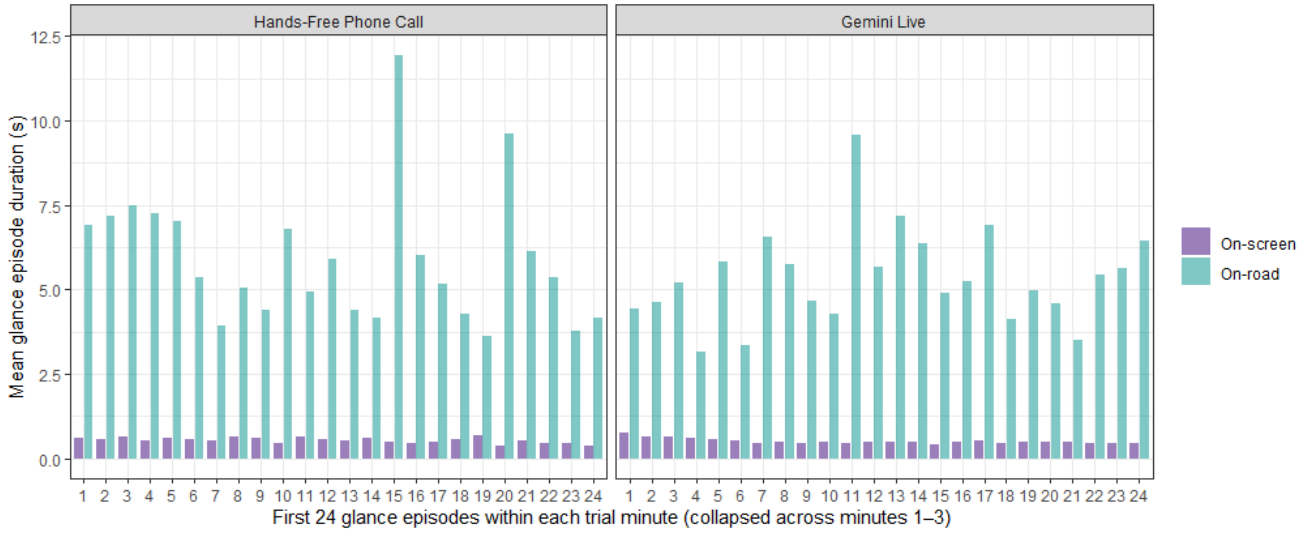


Fig. 3. *Mean glance episode duration as a function of episode order within each trial minute, collapsed across minutes 1–3.*

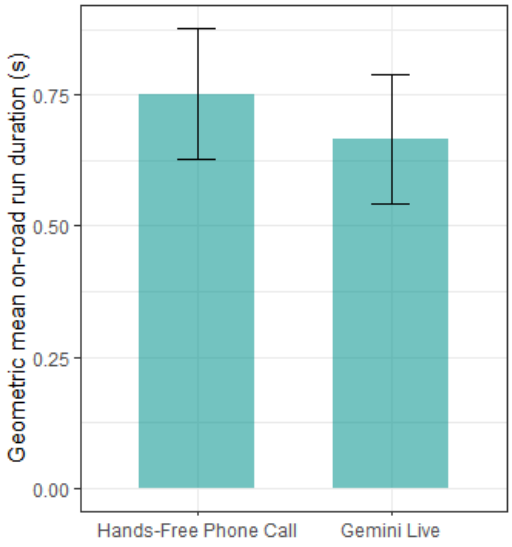


Fig. 4. *Estimated marginal means of on-road run duration by task.*

4. Discussion

This study examined visual attention during hands-free phone conversation and extended, multi-turn conversational AI interaction using a glance-episode-based framework. Rather than relying on aggregate measures of visual demand, the analyses focused on how drivers temporally organize visual attention—specifically, how brief device-directed glances are interleaved with sustained periods of on-road viewing—consistent with prior naturalistic evidence that the timing and structure of glances are critical for safety outcomes (Klauer et al., 2006; Victor et al., 2015).

Across all analyses, on-road glance episodes were substantially longer than on-screen episodes, reflecting the dominance of roadway monitoring during both tasks. However, mean glance episode duration did not differ meaningfully between Hands-Free Phone Call and Gemini Live Multi-turn interaction. Sequential analyses further showed that the ordering and stability of glance episodes over time were similar for both tasks, aligning with prior work indicating that well-designed voice interfaces can produce glance patterns resembling baseline driving rather than visual–manual interaction (Reimer et al., 2021).

At a higher structural level, periods of uninterrupted roadway monitoring between device interactions—operationalized as on-road runs—were also comparable across tasks. These sustained on-road viewing intervals are consistent with prior evidence that several seconds of continuous roadway observation may be sufficient to support the rebuilding of situation awareness following off-road glances (Meyer et al., 2022; Seaman et al., 2021). Despite differences in interaction length and conversational complexity, extended conversational AI interaction did not alter the temporal structure of visual attention relative to a well-established hands-free benchmark.

Taken together, the results suggest that drivers engage in stable, self-regulated visual attention strategies during both hands-free phone conversation and multi-turn conversational AI interaction. Evaluated through the lens of glance episode structure, extended conversational AI interaction did not introduce detectable changes in how drivers allocate visual attention during driving, reinforcing the importance of examining temporal organization in addition to aggregate visual demand metrics emphasized in existing guidelines.

5. Conclusions

This study used a glance-episode-based approach to examine visual attention during hands-free phone conversation and extended, multi-turn conversational AI interaction. Although on-road glance episodes were substantially longer than on-screen episodes across both tasks, mean episode duration, episode sequencing, and on-road run structure did not differ meaningfully between Hands-Free Phone Call and Gemini Live interactions. These findings indicate that extended conversational AI interaction can be supported by visual attention patterns similar to those observed during conventional hands-free phone use, with

brief device-directed glances embedded within sustained periods of roadway monitoring. Evaluating visual behavior at the episode level provides a complementary perspective beyond aggregate visual demand metrics and offers a useful framework for assessing emerging voice-based in-vehicle technologies.

6. Acknowledgments

This work was supported by Google, Inc., Mountain View, CA. The funder played a role in the conceptualization of the study, provided access to Gemini Live integrated with Android Auto, and contributed the decision to submit the manuscript for publication.

References

Alliance of Automobile Manufacturers. (2006). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems*. Washington, DC.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-Car Naturalistic Driving Study data* (DOT HS 810 594). National Highway Traffic Safety Administration.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and the risk of road crashes. *New England Journal of Medicine*, 370(1), 54–59.

Meyer, J., Llaneras, E., & Fitch, G. M. (2022). Impact of temporary browsing restrictions on drivers' situation awareness when interacting with in-vehicle infotainment systems. *Safety*, 8 (4), 81.

Monk, C., Ayala, A., Yu, C. S. P., Fitch, G. M., & Gruber, D. (2026). Visual and cognitive demands of a large language model-powered in-vehicle conversational agent. *arXiv* preprint arXiv:2601.15034.

National Highway Traffic Safety Administration. (2013). *Visual-manual driver distraction guidelines for in-vehicle electronic devices* (78 FR 24818).

Reimer, B., Mehler, B., Muñoz, M., Dobres, J., Kidd, D., & Reagan, I. J. (2021). Patterns in transitions of visual attention during baseline driving and during interaction with visual–manual and voice-based interfaces. *Ergonomics*, 64(11), 1429–1451.

Seaman, S., Seppelt, B., Angell, L., Reimer, B., & Mehler, B. (2021). Vehicle control and response to emerging events: It's both off-road and on-road glance duration. *Proceedings of the 7th International Conference on Driver Distraction and Inattention*, Lyon, France, Virtual, 18–20.

Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D. J., & Biondi, F. (2017). *Visual and cognitive demands of using in-vehicle infotainment systems*. AAA Foundation for Traffic Safety.

Victor, T., Dozza, M., Bärgman, J., Boda, C.-N., Engström, J., Flannagan, C., Lee, J. D., & Markkula, G. (2015). *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk* (SHRP 2 Report S2-S08A-RW-1). Transportation Research Board.

Driver interaction with central display in real-world driving conditions

Mirta Zelenika Zeba [1*], Thomas Streubel [1]

[1] Volvo Cars Cooperation, Safety Center, Göteborg, Sweden
[*]mirta.zelenika.zeba@volvocars.com

Abstract: The present study investigates real-world interactions with the central stack by examining phone-related behaviours, mirror-adjustment practices, and general CD use. Naturalistic driving data from 1,281 vehicles, comprising approximately 575,000 drive cycles recorded between February 2023 and November 2024, were analysed to identify how drivers initiate phone calls. Additional data collected from February to August 2025 captured all mirror-adjustment events, including their timing and driving-state context. Planned analyses further aim to quantify CD interaction frequency and associated visual-attention demands.
Results show that handheld phone use remains the most common method of call initiation (41%), followed by the CD call list (30%), smartphone applications (19%), and speech-based commands (10%). Mirror adjustments occurred predominantly at standstill (64%) and within the first eight minutes of driving, with 44% of drivers not adjusting mirrors during the observation window.

1. Introduction

The implementation of new technologies for in-vehicle interaction, such as in-car touchscreens, provides several functional advantages, including access to multiple features within a single interface and increased convenience for users. However, these technologies also introduce notable drawbacks. In particular, the absence of tactile feedback often requires drivers to devote greater visual attention to the screen compared with physical buttons, thereby increasing distraction and reducing attention to the road environment (Strayer, et al., 2019). Overall, physical buttons located in the centre stack—rather than touchscreen controls—have been associated with reduced driver distraction and improved performance on both secondary tasks and driving-related activities (DeGuzman, Demir, Ayas, & Donmez, 2026) .

Despite these findings, it remains critical to identify which aspects of display design and touchscreen command types most adversely influence driving performance and elevate safety risks. One significant contributor to increased task duration and visual demand is the amount of information presented on the touchscreen (Kujala & Salvucci, 2015). Furthermore, secondary activities—such as adjusting audio settings, modifying in-vehicle configurations, or initiating phone calls—tend to have a less pronounced effect on driving performance than tasks directly tied to vehicle operation. Functions essential to driving, such as activating windshield wipers, have been linked to increased crash risk when executed through touchscreen interfaces (Pichen, Miller, & Baumann, 2021).

Drivers commonly adjust their behaviour when conditions become more demanding. Studies show that during unexpected or critical events, they reallocate attention, slow down, or stop secondary tasks, shifting more glance time to the forward roadway (Metz, Schömig, & Krüger, 2011).

Drivers also anticipate upcoming demands by gathering information before looking away, adapting attention based on the expected duration of eyes-off-road glances and the driving context (Liu, Yuan, & Ma, 2021).

Naturalistic and on-road studies generally indicate that drivers allocate approximately 80–90% of their visual attention to the forward roadway, with the remaining time distributed across mirrors, dashboard instruments, and in-vehicle systems (Victor, Harbluk, & Engström, 2005; Gaspar & Carney, 2019). A landmark naturalistic dataset identified non-forward driver inattention as being present in nearly 80% of crashes and 65% of near-crashes (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). However, such findings likely reflect that distraction often coincides with moments that are already becoming critical, rather than suggesting that brief non-forward glances are inherently unsafe.

Understanding how much of drivers' non-forward visual attention is directed toward infotainment systems—particularly the Central Display (CD)—is therefore essential. While on a situational level time on task is of high relevance, the total-eyes-off-road time (TEORT) is a metric for unspecific driving on a global scale. Estimating the proportion of off-road glances attributable to CD interactions is a crucial step toward identifying the conditions under which such interactions meaningfully influence driving safety. Against this background, the present research aims to examine patterns of CD use, including phone-related interactions, mirror-adjustment behaviour, and the overall frequency of engagement with the central stack, and to quantify the associated visual-attention demands.
Based on these objectives, the study addresses the following research questions:
To what extent do drivers rely on the Central Display (CD) for phone-related tasks instead of interacting with handheld devices during real-world driving?
How frequently are mirror adjustments performed during driving compared with standstill conditions?
How frequently do drivers interact with the Central Display while driving, and under which driving conditions do these interactions occur?

2. Method

2.1 Call initiation

Data related to phone-call initiation were collected between February 2023 and November 2024.

The dataset includes 1,281 vehicles, resulting in approximately 575 thousand recorded drive cycles across Sweden (mostly in the area around Gothenburg). No experimental instructions were provided to drivers; the data therefore represent genuine, everyday driving behaviour. Vehicle signals indicating a call initiation were collected to distinguish between calls made through the handheld phone (phone connected to the vehicle), using the centre display call list, through Apps (e.g., via mirroring or connected apps) and using speech-based voice commands.

2.2 Mirror adjustments

Mirror-adjustment behaviour was analysed using data collected from February 2025 to August 2025. The dataset includes measurements from 164 vehicles running approximately 42 thousand drive cycles in Sweden (mostly in the area around Gothenburg). Only vehicles were included with mirror adjustments through the central display. The extracted variables included the occurrence of mirror-adjustment events, the vehicle's motion state (standstill versus moving), the timing of each adjustment relative to the start of the drive cycle, and the driver-level frequency of mirror adjustments across the observation period.

3. Planned analysis

Beyond the phone-specific and mirror-adjustment analyses, the broader aim of the study is to investigate general interactions with the Central Display (CD). The planned analyses include quantifying the frequency of Central Display (CD) interactions on a time-base (contrary to drive cycle base), estimating the visual-attention demands associated with CD-related tasks, examining how CD use varies across different driving conditions (e.g., standstill vs. moving), and assessing the extent to which CD interactions contribute to non-forward visual attention.

4. Results

Analysis of phone-related behaviour indicates that drivers rely on a variety of methods to initiate calls while driving. Handheld phone use remained the most common method, accounting for 41% of call initiations, despite legal implications and safety concerns related to phone usage while driving. The Central Display (CD) call list was used in 30% of call initiations, demonstrating a moderate but not dominant shift toward in-vehicle phone integration. Additionally, 19% of calls were initiated via smartphone applications accessed through the CD or device mirroring, while 10% were initiated using speech-based voice control (see Figure 1).

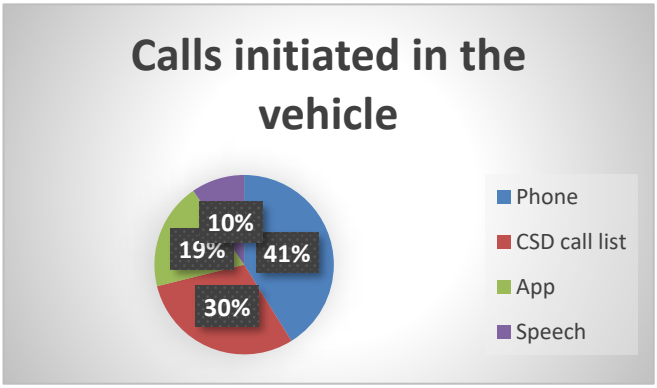


Figure 1. Call initiation in vehicle through different methods

Mirror-adjustment activity during driving was rare. In 44% of vehicles no adjustment was recorded in the measurement period, suggesting that mirror settings are often configured at first vehicle setup. The majority of adjustments—63%—occurred while the vehicle was stationary. A significant portion of adjustments during driving were performed at speeds below 20 km/h. Furthermore, half of the mirror adjustments were within the first 8 minutes of a drive cycle. Overall, events were very rare with less than 1% of drive cycles containing any mirror adjustments with common frequencies of one or two adjustments per drive cycle.

5. Discussion

The findings of this study highlight several important insights into how frequent drivers interact with in-vehicle systems, particularly in the context of phone use, mirror adjustments, and overall reliance on the Central Display (CD). A key result concerns the methods by which drivers initiate phone calls. Although nearly half of all calls were initiated using safer, built-in hands-free options—specifically the central display (CD) call list and speech-based commands—the fact that handheld phone use still accounted for 41% of call initiations highlights a persistent gap between the availability of safer CD-based alternatives and actual driver behaviour.

The mirror-adjustment findings contribute a different dimension to the discussion. Mirror changes were rare during active driving and predominantly occurred when the vehicle was stationary, with most adjustments taking place within the first few minutes of a trip. These patterns suggest that mirror-adjustment behavior is strongly tied to initial vehicle-setup routines rather than ongoing driving demands. However, an important limitation of the dataset is its restricted time window, which does not represent the full lifecycle of vehicle ownership. Nevertheless, the fact that mirror controls are seldom used while driving reinforces the need for these essential safety-related interfaces to remain easily accessible, simple to operate, and designed to minimize any need to divert attention from the forward roadway.

Taken together, these findings underscore the broader challenge of aligning human behavior with the safety expectations of vehicle interface design. While the CD consolidates multiple functions into a single access point, its effectiveness depends heavily on user experience and task efficiency. When built-in systems do not fully meet user needs—or when drivers perceive alternative methods as faster or more familiar—they may revert to behaviors that increase visual distraction and reduce compliance with legal and safety standards.

6. Conclusions

This study examined how drivers initiate phone calls, adjust mirrors, and planned analysis includes interaction with the Central Display (CD) during real-world driving. The results show that handheld phone use remains common indicating that current in-vehicle calling features have not replaced handheld interactions. Mirror adjustments occurred primarily when the vehicle was stationary and early in the drive cycle, suggesting that such tasks are largely tied to initial setup rather than ongoing driving.

References

DeGuzman, C. A., Demir, B., Ayas, S., & Donmez, B. (2026). A review of in-vehicle touchscreen safety and usability,. *Transportation Research Interdisciplinary Perspectives,*, 101873.

Gaspar, J., & Carney, C. (2019). The Effect of Partial Automation on Driver Attention: A Naturalistic Driving Study. *Human factors, 61(8),*, 1261–1276.

IET Digital Library. (n.d.). *Author Guide - IET Research Journals*. Retrieved November 27, 2014, from IET Digital Library: http://digital-library.theiet.org/journals/author-guide

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. *Virginia Tech Transportation Institute. United States. Department of Transportation. Office of Human-Vehicle Performance Research*, DOT HS 810 594.

Kujala, T., & Salvucci, D. (2015). Modeling visual sampling on in-car displays: The challenge of predicting safety-critical lapses of control,. *International Journal of Human-Computer Studies, Volume 79,*, 66-78.

Liu, Z., Yuan, W., & Ma, Y. (2021). Drivers' Attention Strategies before Eyes-off-Road in Different Traffic Scenarios: Adaptation and Anticipation. *International journal of environmental research and public health, 18(7),*, 3716.

Metz, B., Schömig, N., & Krüger, H.-P. (2011). Attention during visual secondary tasks in driving: Adaptation to the demands of the driving task. *Transportation Research Part F: Traffic Psychology and Behaviour, Volume 14, Issue 5,*, 369-380.

Pichen, J., Miller, L., & Baumann, M. (2021). Cooperative Speed Regulation in Automated Vehicles: A Comparison Between a Touch, Pedal, and Button Interface as the Input Modality. *IEEE Intelligent Vehicles Symposium (IV), Nagoya, Japan*, 245-250.

Strayer, D. L., Cooper, J. M., McCarty, M. M., Getty, D. J., Wheatley, C. L., Motzkus, C. J., . . . & Horrey, W. J. (2019). Visual and cognitive demands of CarPlay, Android Auto, and five native infotainment systems. *Hum. Factors, 61 (8)*, 1371-1386.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty,. *Transportation Research Part F: Traffic Psychology and Behaviour, Volume 8, Issue 2,*, 167-190.

Behaviour of cognitively distracted cyclists in T-junctions

Mette Møller[1*], Frauke Luise Berghoefer[2], Mark Vollrath[2]

[1] Technical University of Denmark (DTU), Division of Transportation Science, Akademivej 358, 2800 Kgs. Lyngby, Denmark
[2] Technische Universität Braunschweig, Department of Traffic and Engineering Psychology, Gaußstr. 23 38108 Braunschweig, Germany
[*]mette@dtu.dk

Abstract: This study examines the behavioral effects of hands-free cognitive secondary task engagement on cyclists navigating complex interactive traffic situations. Using a high-fidelity bicycle simulator, 46 participants were randomly assigned to either a control group or a test group performing an acoustic speech task. Participants navigated two randomized scenarios involving different behaviors of the car with regard to the right-of-way and two sight conditions (free vs. restricted). Key metrics included mean speed and braking distance when approaching the intersection. Results indicated no significant main effect of the secondary task on overall behavior. However, significant interactions were found between the speech task and car behavior. Control participants dynamically adapted their braking and speed to specific situational demands (e.g., braking later when a car had already cleared an intersection). Conversely, distracted cyclists exhibited a lack of situational adaptation, failing to increase speed during good sight conditions and braking earlier/more cautiously regardless of whether the car yielded. The findings suggest that while cognitive distraction did not lead to immediate danger in these scenarios, it impaired the participants' ability to fully and rapidly process environmental cues. This led to less efficient, albeit more cautious, behavior. Future research should investigate whether this lack of adaptation translates into increased risk in time-critical or more hazardous traffic environments.

1. Introduction

Previous results on effects of cognitive secondary task engagement among cyclists are mixed. Some studies have found detrimental effects such as failing to slow down when entering an intersection (e.g., de Waard et al., 2014) or increased engagement in violations associated with near-crashes (e.g. Useche et al., 2024). Other studies have found positive compensating effects such as reduced riding speed (e.g., de Waard et al., 2014; Kircher et al., 2015) and limiting secondary task engagement to less demanding situations (e.g., Brandt et al., 2021; Kircher et al., 2015).

With increasing possibility to engage in hands-free secondary tasks using headphones, the behavioural effects of such engagement should be further examined. Results of a recent study using simple traffic situations (Møller et al., 2024) found the effect of cognitive secondary task engagement to depend on the complexity of the cycling task. Therefore, this study aims to examine the behavioural effects of handsfree cognitive secondary task engagement during complex interactive traffic situations. The study is theoretically based on Wickens' Multiple Resource Theory (Wickens, 2002) implying in this combination that interferences could happen at a level of cognitive processing and decision making.

2. Method

2.1 Bicycle simulator

Data was collected using a bicycle simulator (see Fig. 1) at TU-Braunschweig, Germany. The bike is mounted on a static platform allowing side tilts. Monitors enable a 360° view, and a fan provides a headwind depending on the cycling speed. Bluetooth headphones provide ambient sounds (e.g., passing vehicles). The simulator runs with the software SILAB 7.0 (Würzburg Institute for Traffic Sciences, 2022).



Fig. 1. *The bicycle simulator used in the study*

2.2 Secondary task

The study included two secondary task conditions: a control group with no task (control group) and a test group with an acoustic speech task (speech task). For the speech task, participants used Bluetooth headphones to listen to three color names and a rule for reordering them. They then vocalized the new order. This involved acoustic input, visual-spatial representation, cognitive processing and decision and speech output (see Vollrath & Totzke, 2005).

2.3 Experimental design and parameters

Two basic scenarios were used (see Fig. 2). The cyclist was riding straight and had the right of way. A car war either going on the road beside the cyclist and then turning right or coming from the side road to turn right. In both scenarios, the cyclist had the right of way.

The scenarios were varied regarding car behaviour (rule compliant yes vs no), and sight condition (free vs restricted). In total the study included eight different scenarios which where experienced by all cyclists (see Fig 3). The order of the scenarios was randomized.

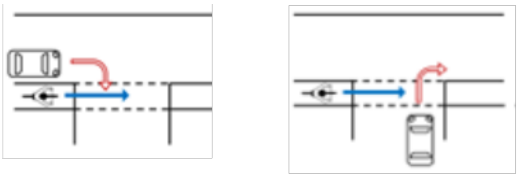
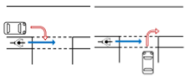


Fig. 2. *The two basic scenarios.*



		restricted sight	free sight of intersection
car waits	control	P01 .. P23	P01 .. P23
	speech task	P24 .. P46	P24 .. P46
car drives through	control	P01 .. P23	P01 .. P23
	speech task	P24 .. P46	P24 .. P46

Fig. 3. *Experimental design ("P": participant) with four independent groups*

To analyze cyclists' behavior at the intersection, we defined a crossing point of car and cyclist trajectories. We studied behavior from 10 meters before to 10 meters after this point for passing, and from 90 meters before to 10 meters before for approaching. Key metrics included mean and minimum speeds, braking distance, maximum braking force, and SDLP for lateral control. During intersection negotiation, we examined the duration and mean speed to assess handling speed, along with SDLP for lateral control.

2.4 Procedure

After agreeing to participate and data privacy, participants were randomly assigned to one of the two secondary task conditions and introduced to the bicycle simulator. They practiced turning and braking on a test course and handling traffic in a city ride until comfortable. The test ride lasted about 30 minutes. Afterwards, participants filled out a demographic questionnaire, received thanks, and were compensated. The entire experiment took about 1 hour.

2.5 Participants

Participants were recruited via mailing lists and online forums at the TU-Braunschweig. Participants should be able to ride a bike and know the German traffic rules. They received € 5 per half hour or credit points. The study was approved by the ethical commission at the Institute of Psychology at the TU- Braunschweig.

The sample included 46 participants (15 female, 31 male) aged 21 - 76 years (M = 38.3 years, SD = 15.9). Most participants were employees (44%) or students (41%), and the majority cycled daily (52%) or almost daily (ca. 28%).

2.6 Data analysis

To examine the effects a mixed 2 (secondary task, between) x 2 (car behavior, between) x 2 (sight condition, within) factor ANOVA was conducted for each of the two basic scenarios. The statistic software IBM SPSS Version 28 was used.

3. Results

3.1 Scenario 1: Cyclist rides straight – car turns right to side road

When approaching, aan interaction was found between car behavior and secondary task for the point when cyclists started to brake ($F(1,44)=5.9$, $p = 0.019$; see Fig. 4). In the control group, cyclists braked earlier when the car drove at their side and then stopped. When it overtook them and turned right, they braked quite late, because mostly the car was already gone when they reached the intersection. The speech task group braked a little later when the car drove beside them and stopped, but also did that although the car overtook them and turned right.

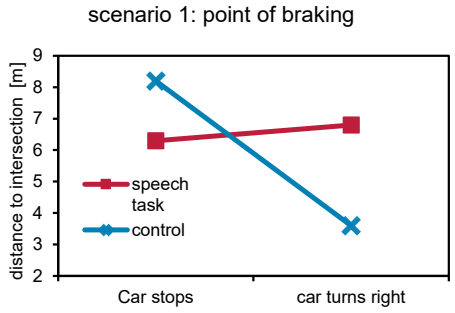


Fig. 4 Average distance to the intersection when the car stopped as compared to when it turned right in the two groups in the first scenario where the car drove along the cyclists

Regarding sight, there was an interaction between sight and secondary task was found for mean speed ($F(1,44)=5.9$, $p=0.019$; see Fig. 5). The control group adapted their speed by going slower at restricted sight. The speech task group did not show this adaptation but cycled more slowly even with good sight.

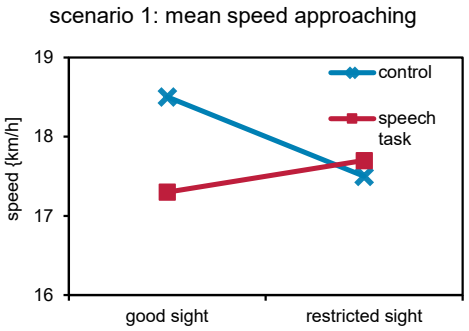


Fig. 5 *Mean speed when approaching the intersection with good sight and restricted sight in the two groups in the first scenario where the car drove along the cyclists*

In the intersection, there was no effect of secondary task.

3.2 Scenario 2: Cyclist rides straight – car comes from the right

As in the first scenario, there is a tendency of an interaction between group and speech task for the point of braking when approaching the intersection ($F(1,44)=3.9$, $p=0.055$).

In the control group drivers brake later when the car stops and lets them pass (Fig 6). When the car takes the right of way and turns right they brake earlier to avoid a collision. Drivers in the speech task group brake similarly early when the car turns right. However, when the car stops, they brake even earlier.

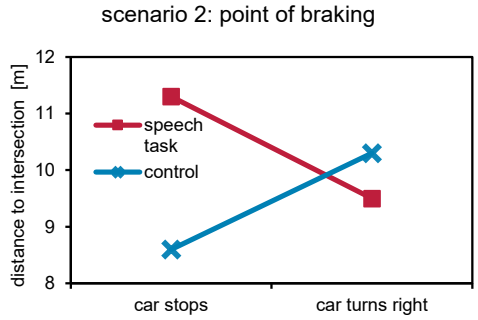


Fig. 6 *Average distance to the intersection when the car stopped as compared to when it turned right in the two groups in the first scenario where the car drove along the cyclists*

4. Discussion

Overall, there was no main effect of the acustic speech task on the behavior of the cyclists in these two complex scenarios in any of the parameters examined.

However, in both scenarios there was an interaction between speech task and the behavior of the car (giving the right of way and stopping vs. taking the right of way and turning right before the cyclist). Cyclists in the control group adapted their behavior to this variation and braked later in the first scenario when the car turned right before them, because the car was already gone when they reached the intersection. They braked earlier in the second scenario when the car turned right in order to avoid getting too close to this car when turning.

With the acoustic speech task, cyclists did not brake as early in the first scenario as the control group, when the car stopped. Moreover, they did not brake later when the car was already gone because it had turned right before them. Both reactions might be an indicator that with the acoustic speech task the situation is understood a bit later than in the control group. However, the slower reaction when the car stopped did not leed to any dangerous situations. Braking earlier when the car had turned is a more cautious behavior, perhaps also because not quite understanding what was going to happen.

Similarly, when looking at the mean speed when approaching cyclists with the acoustic speech task failed to adapt their speed to the sight condition as the control group did by going faster with good sight. Thus, with the speech task and good sight cyclists went more slowly, perhaps as a compensation for the resources required for this task.

While there is no negative or dangerous effect of cognitive distraction on cycling even in these more complex situations, cognitive distraction seems to impair the full and fast understanding of relevant aspects of the driving situation like the behavior of the car (yielding or taking the right of was) or the sight towards car traffic (free vs. restricted sight). This is in line with what we expected from the Multiple Ressource Theory (Wickens, 2022). It also indicates a lack of adaptation to the overall situation as found in other studies (e.g., De Ward, 2015; Kircher et al., 2015).

In the scenarios examined here, this did not leed to crashes or dangerous situations because of the way the scenarios were designed. Here, the lack of adaptation lead to a more cautios behavior: Braking earlier although the car had already turned right and was gone when they reached the intersection in the first scenario, braking earlier although the car had stopped and let them pass in the second scenario and cycling more slowly although there was a good sight towards motor traffic.

5. Conclusions

Cognitively distracted cyclists don't adjust their behaviour to the characteristics of more complex situations. However, the scenarios examined did not lead to dangerous or critical situations. Thus, the next study should focus on situations where not adapting the behavior to the requirements of the situations may lead to critical situations.

6. Acknowledgments

The study was funded by the Danish Accident Investigation Board.

References

Brandt, R. K., Haustein, S., & Møller, M. (2021). Cyclists' phone use in relation to proximate environmental characteristics – A qualitative study. *Journal of Transport & Health*, 23, Article 101283. https://doi.org/10.1016/j.jth.2021.101283

de Waard, D., Lewis-Evans, B., Jlijs, B., Tucha, O., & Brookhuis, K. (2014). The effects of operating a touch screen smartphone and other common activities performance while bicycling on cycling behaviour. *Transportation Research Part F*, 22, 196–206. https://doi.org/10.1016/j.trf.2013.12.003

Kircher, K., Ahlstrom, C., Palmqvist, L., & Adell, E. (2015). Bicyclists' speed adaption strategies when conducting self-paced vs. system-paced smartphone tasks in traffic. *Transportation Research Part F*, 28, 55–64. https://doi.org/10.1016/j.trf.2014.11.006

Møller, M., Berghoefer, F.L., & Vollrath, M. (2024). How does hands-free cognitive distraction influence cycling behaviour and perceived safety? *Transportation Research Part F,* 104, 522-531. https://doi.org/10.1016/j.trf.2024.06.026

Useche, S. A., Alonso, F., Montoro, L., & Esteban, C. (2018). Distraction of cyclists: How does it influence their risky behaviors and traffic crashes? *PeerJ,* 6, e5616

Vollrath, M., & Totzke, I. (2005). Secondary tasks while driving effects and countermeasures. *Advances in Transportation Studies*, 7, 67-80

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3, 159-177. https://doi.org/10.1080/146392202101

Würzburg Institute for Traffic Sciences, 2022.

Age-Dependent Changes in Pupillometric Indicators of Cognitive Load During Hands-Off L2 Driving

Courtney M. Goodridge [1,2*], Rafael C. Gonçalves[2], Gaëtan Merlhiot[3], Michael G. Lenné[4], Jonny Kuo[4] Natasha Merat[2]

[1] University of Leeds, School of Psychology, Leeds, United Kingdom
[2] University of Leeds, Institute for Transport Studies, Leeds, United Kingdom
[3] VEDECOM Institute, Versailles, France
[4] Seeing Machines, Canberra, Australia
[*] c.m.goodridge@leeds.ac.uk

Abstract: Arousal is a physiological and psychological state that is crucial for regulating cognitive control. The Locus Coeruleus-Norepinephrine (LC-NE) system is a brain structure implicated in this regulation; with pupillometry used as a non-invasive indicator of LC-NE activity. Therefore, pupillometry may be a useful tool for evaluating driver cognitive load, using camera-based Driver Monitoring Systems (DMS). With an ageing population and increasing implementation of automated systems in vehicles, it is important to understand how pupillometry fluctuates with age. We conducted a driving simulator experiment measuring drivers' (N = 38) pupil activity during hands-off L2 driving and transitions of control to manual driving, under high and normal cognitive load. Linear multilevel models revealed quadratic relationships between pupillometry and age: pupil activity peaked for middle aged drivers and declined for older drivers. There was also evidence that older-than-average drivers were less affected by cognitive load, as indexed by mean pupil diameter. These findings reiterate the need for DMS algorithms to be trained on representative samples and suggest that DMS profiles may need to adapt dynamically as drivers age.

1. Introduction

As road vehicles become more automated, there is increased emphasis on Driver Monitoring Systems (DMS) predicting Driver Readiness (DR), which is defined as the probability of successful takeovers (ISO, 2025). This is particularly important for intermediate automation levels (e.g., SAE Level 2 [L2]) where lateral and longitudinal dynamics are controlled by the vehicle, yet the driver must continue to monitor the road and be ready to regain manual control of driving. There have been many attempts to create standardised DR metrics (Gonçalves et al, 2024; Kim et al, 2022; Mariajoseph et al, 2020), yet DR remains an amalgamation of varying constructs such as cognitive readiness (i.e., attention and cognitive load), motoric readiness (i.e., motoric availability), and individual factors (i.e., driver experience, age) (Gonçalves et al, 2024).

Cognitive load refers to the amount of information processing resources required for completing a task (Engström et al, 2017). Identifying dangerous levels of cognitive load is one of the biggest challenges for DMS, since peripheral methodologies must be used to predict an individual's internal cognitive state. Pupil size is one of the easiest, non-invasive, peripheral methods for estimating neural activity associated with cognitive effort (Vilotijević & Mathôt 2024). Pupillometry is also linked to the underlying activity of the Locus Coeruleus-Norepinephrine (LC-NE) system which plays a role in regulating global arousal (Mathôt, 2018).

Given that arousal is thought to modulate cognition (Gruseschow et al, 2020), researchers have used mean and standard deviation (SD) of pupil diameter to approximate LC-NE system activity during transitions of control from automated driving (Goodridge et al, 2026; Radhakrishnan et al, 2023). These studies show that pupillometry can provide real-time measures of cognitive processing for future DMS during system monitoring and control transitions. A crucial next step is to investigate pupillometry differences across common demographics - such as age - given a globally ageing driver population. Driver demographics are known to impact driver state estimation (Goodridge et al, 2024; 2026; Gonçalves et al, 2025), and age-related impairments in the LC-NE system are linked with natural cognitive decline (Kremen et al, 2019). We therefore built upon previous research (Goodridge et al, 2026) to investigate age-related differences in pupillometry associated with cognitive load during hands-off L2 driving and transitions of control.

2. Method

2.1 Participants

This study recruited 38 participants (males = 22; females = 16). Average age was 38.81 years (range = 22-65 years old, SD = 12.502). Participants had normal or corrected to normal vision, a UK driving license (mean = 17.84 years, range = 4-43 years, SD = 12.119), and were regular drivers (mean = 9355.26 miles, range = 5000-20000, SD = 4153.97).

2.2 Apparatus

The experiment was conducted in the second-generation University of Leeds Driving Simulator (see Figure 1). Pupillometry data were captured using a Seeing Machines DMS eye tracker, sampling at 60 Hz.



Figure 1: The University of Leeds Driving Simulator.

2.3 Design and procedure

Procedures were approved by the University of Leeds Research Ethics Committee (Reference code: 2022-0353-206) and informed consent was obtained prior to data collection.

A 2 (event criticality: time to collision [TTC] = 3 s; TTC = 5 s) x 2 (cognitive load: 2-back, no 2-back) repeated-measures design was used. Participants completed two UK motorway drives: one with a 2-back task (Mehler et al, 2011) during hands-off L2 driving, and another without an additional task. Each drive included 10 events consisting of 30 s of manual driving followed by two minutes of hands-off L2 driving. A request to intervene (RTI) was then issued via an auditory tone and a steering-wheel icon in the dash area.

Critical events involved a lead vehicle decelerating at the RTI (two severe: TTC = 3 s; two moderate: TTC = 5 s). Non-critical events either had no lead vehicle (two) or a lead vehicle that maintained speed (four).

2.4 Statistical modelling

We used linear multilevel models (MLMs) to analyse pupillary activity. During hands-off L2 driving, we modelled mean and SD of pupil diameter as a function of 2-back, lead vehicle, and age. We also modelled mean pupil diameter during the 10 s transition of control following the auditory tone as a function of 2-back, event criticality, and age. We fitted age as linear and polynomial terms to assess non-linear trends with pupillometry.

3. Results

3.1 Effect of 2-back, lead vehicle, and age on pupil diameter during hands-off L2 driving

Drivers' pupils were larger during 2-back (4.345 mm) relative to no 2-back trials (4.080 mm) (β = 0.284, [0.201, 0.368]). Furthermore, age demonstrated a negative quadratic relationship with mean pupil diameter (β = -5.008, [-9.182, -0.833]). Mean pupil diameter began decreasing significantly from 38.53 years of age. Finally, there was a potential interaction between 2-back and age (β = -1.688, [-3.956, 0.579]), indicating that data were compatible with negligible and modest positive effects. This interaction was driven by 2-back having less impact on mean pupil diameter for older drivers (2-back = 4.069 mm versus no 2-back = 3.860 mm) relative to younger drivers (2-back = 4.205 mm versus no 2-back = 3.845 mm).

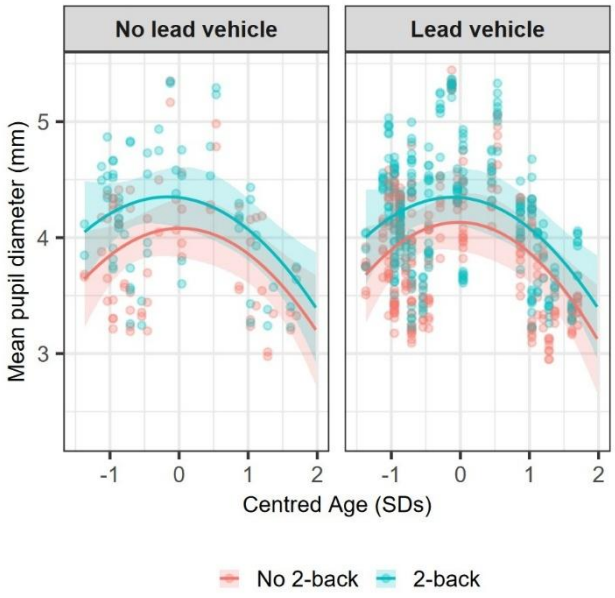


Figure 2: Mean pupil diameter across age, 2-back, and lead vehicle.

SD of pupil diameter was larger during 2-back (0.277 mm) relative to no 2-back trials (0.242 mm) (β = 0.041, [0.022, 0.061]). Age also showed evidence of a negative quadratic relationship with the SD of pupil diameter (β = -0.959, [-1.756, -0.163]); SD of pupil diameter declined from 38.62 years of age.

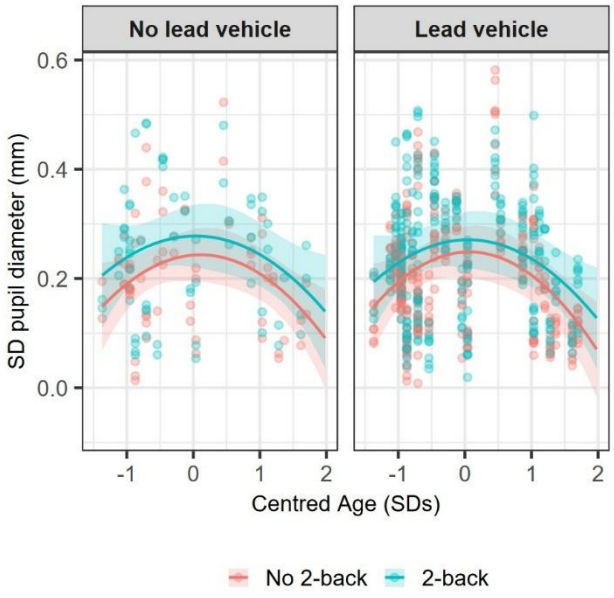


Figure 3: SD of pupil diameter across age, 2-back, and lead vehicle.

3.2 *Effect of 2-back, event criticality, and age on pupil diameter during transitions of control*

There was an interaction between 2-back and event criticality. Drivers' pupils were larger during critical transitions relative to non-critical transitions; however, this effect was more pronounced following no 2-back trials (non-critical = 4.091 mm; TTC 3 s = 4.155 mm; TTC 5 s = 4.173 mm) relative to 2-back trials (non-critical = 4.345 mm; TTC 3 s = 4.348 mm; TTC 5 s = 4.363 mm) (β = 0.063, [0.024, 0.105]). Age demonstrated a negative quadratic relationship with mean pupil diameter, regardless of 2-back or event criticality (β = -5.269, [-9.416, -1.123]).

4. Discussion

Pupillary activity was higher during high cognitive load and critical transitions of control indicating increased effort attributed to non-driving related task engagement and responses to critical events. Furthermore, we found evidence for an inverted U-shaped relationship between pupillometry and age. Middle aged individuals (38.53 years of age) showed the highest pupillary activity, indicating stronger arousal mobilisation. Older-than-average drivers demonstrated diminished pupillary dynamics and showed evidence of pupils being less reactive to 2-back-induced cognitive load.

These data are consistent with age-related changes in pupillary activity that could be due declining LC noradrenergic neurons or the degradation of the pupillary muscles (Lenain et al, 2026). These findings could indicate how natural ageing impacts the external substrate we use to measure driver state (e.g., pupillary muscles); as well as the internal substrate that causes changes in cognition (e.g., neurons). Crucially, we observed these effects during a dynamic applied task without using wearables or fixed-head positions.

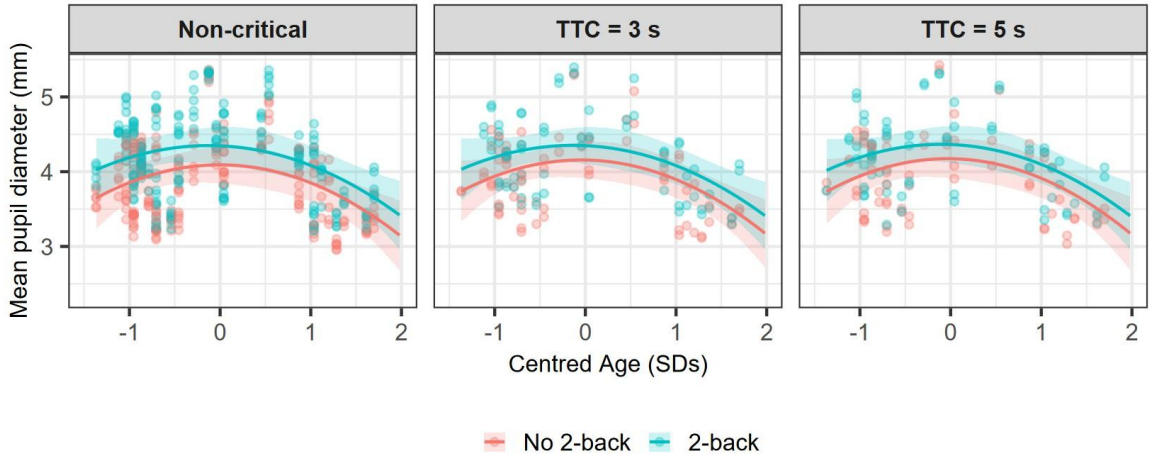


Figure 4: Mean pupil diameter across age, 2-back, and event criticality.

These results reiterate the importance of demographics for improving driver state estimation. Age-related changes in the LC-NE system are known to contribute towards declines in attention and memory (Mather, 2020), suggesting the observed effects could reflect age-associated differences in attentional capacity. Consequently, DMS algorithms should be trained on representative data and may benefit from user-specific pretrained models that adapt dynamically with age.

Secondly, this analysis signals the need for investigating age as a continuous factor. Studies modelling age-related differences in manual and automated driving

(Öztürk et al, 2023) bin age into categories which may miss non-linear trends associated natural cognitive decline.

5. Conclusions

Age demonstrates non-linear relationships with pupillary activity. This could demonstrate age-related LC-NE system changes that alter arousal mobilisation whilst monitoring and transitioning from automated driving systems.

6. Acknowledgments

This research was conducted as part of the Hi-Drive project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101006664.

References

Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human factors*, *59*(5), 734-764.

Gonçalves, R. C., Goodridge, C. M., Kuo, J., Lenné, M. G., & Merat, N. (2024). Using driver monitoring to estimate readiness in automation: a conceptual model based on simulator experimental data. *Cognition, Technology & Work*, *26*(4), 569-584.

Gonçalves, R. C., Pardo, J., Elhenawy, M. M. Z., Kuo, J., Azarmi, M., Rezaei, M., ... & Merat, N. (2025, September). It Matters Who Is Behind The Wheel: Driver Monitoring Feature Analysis Using Explainable AI. In *Adjunct Proceedings of the 17th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 245-249).

Goodridge, C. M., Goncalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., ... & Merat, N. (2024). Gaze entropy metrics for mental workload estimation are heterogenous during hands-off level 2 automation. *Accident Analysis & Prevention*, *202*, 107560.

Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y., … Merat, N. (2026, February 25). Applying Adaptive Gain Theory to Pupillary Dynamics During Hands-Off L2 Driving Under High Cognitive Load. Retrieved from osf.io/preprints/psyarxiv/xcw8r_v2

Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., ... & Merat, N. (2024). The Impact of N-back-Induced Mental Workload and Time Budget on Takeover Performance.

Grueschow, M., Kleim, B., & Ruff, C. C. (2020). Role of the locus coeruleus arousal system in cognitive control. *Journal of Neuroendocrinology*, *32*(12), e12890.

International Organization for Standardization. (2025). *Road vehicles—Driver readiness and intervention management—Part 1: Partial automation (Level 2) (ISO/TS 5283-1:2025)*. https://www.iso.org/standard/85035.html

Kim, J., Kim, W., Kim, H. S., Lee, S. J., Kwon, O. C., & Yoon, D. (2022). A novel study on subjective driver readiness in terms of non-driving related tasks and take-over performance. *ICT Express*, *8*(1), 91-96.

Kremen, W. S., Panizzon, M. S., Elman, J. A., Granholm, E. L., Andreassen, O. A., Dale, A. M., ... & Franz, C. E. (2019). Pupillary dilation responses as a midlife indicator of risk for Alzheimer's disease: association with Alzheimer's disease polygenic risk. *Neurobiology of aging*, *83*, 114-121.

Lenain, M., Sparrow, L., Ott, L., & Coello, Y. (2026). Impact of Cognitive Load on Pupil Dilation: A Comparative Study of Young and Older Adults. *Biological Psychology*, 109210.

Mariajoseph, M., Gallina, B., Carli, M., & Bibbo, D. (2020, May). A physiology-based driver readiness estimation model for tuning ISO 26262 controllability. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)* (pp. 1-5). IEEE.

Mather, M. (2020). The locus coeruleus-norepinephrine system role in cognition and how it changes with aging. *The cognitive neurosciences*, 91-104.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of cognition*, *1*(1), 16.

Mehler, B., Reimer, B., & Dusek, J. (2011). MIT AgeLab delayed digit recall task (n-back). *MIT Center for Transportation & Logistics Research Paper*, (2011/018).

Öztürk, İ., Merat, N., Rowe, R., & Fotios, S. (2023). The effect of cognitive load on Detection-Response Task (DRT) performance during day-and night-time driving: A driving simulator study with young and older drivers. *Transportation research part F: traffic psychology and behaviour*, *97*, 155-169.

Radhakrishnan, V., Louw, T., Gonçalves, R. C., Torrao, G., Lenné, M. G., & Merat, N. (2023). Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving. *Transportation research part F: traffic psychology and behaviour*, *94*, 254-267.

Vilotijević, A., & Mathôt, S. (2024). Functional benefits of cognitively driven pupil-size changes. *Wiley Interdisciplinary Reviews: Cognitive Science*, *15*(3), e1672.

Crash Risk of Eyes off Road among Teen Drivers across Age and Driving Experience

Lesheng Hua [1], Shu Han [1], Allison E. Curry [2], and Sheila G. Klauer[1*]

[1] Virginia Tech Transportation Institute, Division of Vehicle, Driver, and Safety Systems, 3500 Transportation Research Plaza, Blacksburg, VA, USA
[2] Children's Hospital of Philadelphia, University of Pennsylvania, 3401 Civic Center Boulevard, Philadelphia, PA, USA
[*] cklauer@vtti.vt.edu

Abstract: Fatal crashes involving teen drivers remain a major safety concern, mainly due to inexperience and increased susceptibility to distraction. Eyes-off-road (EOR) is a crucial mechanism for assessing visual distraction. This study used two naturalistic driving datasets to examine the relationship between EOR duration and crash/near crash (CNC) risk among teen drivers across two driving phases and two age groups. Odds ratios (ORs) were calculated for both total EOR and single longest EOR at various duration thresholds. Results showed that during the first six months after licensure, ORs increased monotonically from approximately 2 to 11 as total EOR durations increased from one to four seconds. In contrast, in the second six months, OR did not increase monotonically and changed slightly. Similarly, ORs in the 16-17 age group increased monotonically from nearly 2 to 11 as the total EOR durations increased from one to five seconds, whereas ORs in the 18-19 age group remained stable. For the single longest EOR, ORs were significant when the duration exceeded one second in the first six months and two seconds in the second months. In the 16-17 age group, ORs surged from approximately 2 to 45 when durations of single longest EOR increased from one to four seconds, while ORs in the 18-19 age group increased from approximately 2 to 25. Overall, crash risk increased with EOR duration and declined as drivers gain age and driving experience.

1. Introduction

Fatal crashes involving teen drivers remain a major transportation safety concern, mainly due to inexperience, immaturity, and secondary task engagement (Simons-Morton et al., 2015). Crash risk is highest immediately after licensure, particularly during the first several months of independent driving, and declines as drivers gain both age and driving experience (Lee et al., 2011; Simons-Morton et al., 2017).

Secondary task engagement can divert attention from the driving task, leading to inattention to the roadway. One key indicator of this inattention is eyes off road (EOR). Prior research has shown that a single long glance exceeding two seconds can increase crash risk by approximately fourfold among teen drivers (Klauer et al., 2017; Simons-Morton et al., 2014). Younger drivers (often aged 16–17 years) have a higher prevalence of secondary task engagement and are more likely to exhibit longer EOR durations (Simons-Morton et al., 2017; Guo et al., 2016).

Naturalistic driving studies, such as the Second Strategic Highway Research Program Naturalistic Driving Study (SHRP 2 NDS) and the Supervised Practice Driving Study (SPDS NDS), provide detailed observations of real-world driving behaviours among drivers of different ages, including novice teen drivers. These two datasets have been widely used to investigate the relationship between secondary tasks, eye-glance behaviour, and crash risk (Gershon, et al, 2019; Klauer et al., 2014; Guo et al., 2016). Notably, although previous studies have associated crash risk with EOR, age, and driving experience, few studies have directly examined the crash risk of EOR among teen drivers with different ages and driving experience.

2. Method

This study examined the association between crash risk and various off-road glance durations among teen drivers of different ages and driving phases using mixed-effects logistic regression models.

2.1 Two Datasets

The SHRP2 NDS included a subset of 254 drivers aged 16-17 and an additional 234 teen drivers aged 18-19 years. These teen drivers drove for up to 24 months, accumulating approximately 1,800,000 miles, and were involved in 149 crashes.

The SPDS NDS included a subset of 82 novice teenage driver (mean age = 15.6 years at recruitment) with data collected during their first 12 months of independent driving. During this period, the teenagers drove approximately 380,000 miles and were involved in 60 crashes.

2.2 Variables

In the regression model, the response variable was the safety outcome (crash or no crash), while the explanatory variables included age group, driving phases, and eye glance durations. Teen drivers were categorized into two age groups: 16-17 years and 18-19 years. Driving phases in the first year of independent driving after licensure were divided into two phases: first six months of independent driving (months 1-6) and second six months of independent driving (months 7-12).

The predictor of interest was the duration of eyes off road. Two EOR metrics were evaluated: (1) single longest glance of eyes off the roadway, defined as the duration of the single longest glance away from the forward roadway during the 6-second window prior to a crash/near crash (CNC) and during control road segments; and (2) total duration of eyes

away from the roadway, defined as the cumulative time during which the driver's eyes were not on the forward roadway during the 6-second windows prior to each CNC and during the 6 seconds of each baseline. Five duration thresholds (e.g., 1 s, 2 s, 3 s, 4 s, 5 s) were used to categorize all data into two duration groups for each threshold (e.g., ≥ 1 s vs. < 1 s, ≥ 2 s vs. < 2 s, etc.), with the shorter EOR duration group serving as the reference group (i.e., < 1 s, < 2 s, etc.). Separate regression models were fitted for each duration threshold to estimate odds ratios (ORs) for both total EOR and single longest EOR.

3. Results

3.1 Relationship between the Crash Risk of Various Durations and Driving Phases

In the first six months of independent driving, the odds of CNC were 2.51 times higher when total EOR was longer than one second compared with durations shorter than one second. This increased monotonically as eyeglance duration increased, reaching approximately 11 times higher when total EOR exceeded four seconds (see Table 1). In other words, the CNC likelihood increased consistently as glance duration increased from one to four seconds. Similarly, the OR for the single longest EOR exceeding one second was significant and greater than that for total EOR (see Table 2).

In the second six months, the statistical significance of OR shifted from the one-second threshold observed in the first six months to the two-second threshold for both total EOR and single longest EOR. The ORs for other duration thresholds were not significant and did not increase monotonically with glance durations of total EOR.

Compared with the first six months, ORs were lower during the second six months. When the single longest EOR exceeded one second during the first six months, teen drivers may have perceived the situation as risky and quickly returned their gaze to the roadway, resulting in relatively few observations of EOR durations exceeding two seconds. By the second six months, however, teen drivers had gained more driving experience, allowing more instances of longer EOR durations (e.g., exceeding two seconds) to be observed.

3.2 Crash Risk of Various Glance Durations Changes with Age

In the 16-17 age group, the odds of CNC were 1.89 times higher when the total EOR exceeded one second compared with durations shorter than one second (see Table 1). The odds increased monotonically as glance duration increased, reaching 11 times higher when total EOR exceeded five seconds. For both total EOR and single longest EOR, the ORs across various duration thresholds were significant. For the single longest EOR, crash risk surged with each additional second of glance duration, reaching elevenfold when the duration exceeded two seconds and approximately forty-fivefold as the duration reached four seconds (see Table 2).

In the 18-19 age group, the odds of a CNC were 1.66 times higher when total EOR was longer than one second compared with durations shorter than one second. However, the increase of ORs was not monotonic and remained relatively stable across the duration thresholds. The likelihood of a CNC was 25.36 times higher when the single longest EOR exceeded three seconds compared with durations shorter than three seconds.

Compared with the 16-17 age group, ORs across duration thresholds were generally lower for both total EOR and single longest EOR in the 18-19 age group.

4. Discussion

4.1 Driving Phases and Eye Glance Metrics

In the first six months, the significant ORs indicated that crash risk increased with durations of total EOR. In the second six months, the two- second threshold served as a critical point at which the crash risk increased significantly for total EOR and single longest EOR longer than two seconds. A single longest EOR exceeding two seconds had nearly a fivefold increase in crash risk (OR = 4.96), whereas total EOR exceeding two seconds also increased crash risk

Table 1 *OR of Total EOR by driving phases and age groups*

Total EOR	First 6 months OR (95% CI)	Second 6 months OR (95% CI)	16-17 age group OR (95% CI)	18-19 age group OR (95% CI)
≥ 1 s	2.51 (1.29, 4.89) **	1.51 (0.72, 3.19)	1.89 (1.89, 1.90) **	1.66 (1.29, 2.12) **
≥ 2 s	3.59 (1.46, 8.81) **	2.91 (1.08, 7.83) *	3.05 (3.05, 3.05) **	2.13 (1.56, 2.91) **
≥ 3 s	5.71 (1.87, 17.41) **	1.75 (0.43, 7.09)	4.36 (2.54, 7.50) **	2.09 (1.39, 3.14) **
≥ 4 s	11.65(1.96, 69.13) **	1.49 (0.14, 16.21)	5.89 (2.78, 12.49) **	1.95 (1.11, 3.42) *
≥ 5 s			11.00 (10.99, 11.02) **	2.96 (0.94, 9.38) *

Table 2 *OR of Single longest EOR by driving phases and age groups*

Single longest EOR	First 6 months OR (95% CI)	Second 6 months OR (95% CI)	16-17 age group OR (95% CI)	18-19 age group OR (95% CI)
≥ 1 s	2.84 (1.29, 6.20) **	1.99 (0.86, 4.57)	2.20 (2.20, 2.20) **	1.74 (1.33, 2.29) **
≥ 2 s		4.96 (1.01, 24.24) *	11.03 (11.02, 11.04) **	2.47 (1.41, 4.33) **
≥ 3 s			29.89 (5.57, 160.53) **	25.36 (5.02, 127) **
≥ 4 s			45.38 (45.32, 45.43) **	
≥ 5 s			17.51 (1.60, 191.34) **	

*Note: OR = odds ratio; CI = confidence interval; **p< .01; * p<.05*

but to a less extent (OR = 2.91). This difference is expected given previous research suggesting that two second eye-glance metrics increase crash risk. For the total EOR, drivers may have brief glances to the forward roadway and gain some, albeit insufficient, information about the driving environment. In contrast, a single longest EOR is one continuous glance away from the forward roadway, during which the driver receives no visual information about the roadway ahead. Thus, the crash risk of a single longest EOR is much greater than that of total EOR at the same duration threshold.

4.2 Age

Teen drivers aged 18-19 had more single prolonged glances exceeding one second and longer total EOR glances than drivers aged 16-17. Despite this, the crash risk remained higher in the 16-17 age group, suggesting that 18–19-year-old teens may have learned to be more judicious about when to look away. The significance of the ORs for both age groups and both EOR metrics indicates that crash risk increases with EOR duration but declines as teen drivers become older, which aligns with findings from previous studies. A single prolonged glance away from the road is a strong predictor of crash risk.

5. Conclusions

Crash or near-crash (CNC) risk was higher among younger teen drivers aged 16–17 than among those aged 18–19, declined as driving experience increased during the first year of independent driving, and increased with eyes-off-road durations. Especially, single prolonged glances occurred less frequently than total EOR. These findings highlight the importance of improving attention management among novice teen drivers.

6. Acknowledgments

The research reported herein was performed under BTSCRP Project 28 by the Virginia Tech Transportation Institute (VTTI) with support from the Children's Hospital of Philadelphia (CHOP).

References

Lee, S. E., Simons-Morton, B. G., Klauer, S. E., Ouimet, M. C., & Dingus, T. A. (2011). Naturalistic assessment of novice teenage crash experience. *Accident Analysis & Prevention*, 43(4), 1472-1479.

Gershon, P., Sita, K. R., Zhu, C., Ehsani, Z. P., Klauer, S. G., Dingus, T. A. & Simons-Morton, B. G. (2019) "Distracted Driving, Visual Inattention, and Crash Risk Among Teenage Drivers," American Journal of Preventive Medicine, 56(4): 494-500. https://doi.org/10.1016/j.amepre.2018.11.024

Guo, F., Klauer, S. G., Fang, Y., Hankey, J. M., Antin, J. F., Perez, M. A., Lee, S. E., & Dingus, T. A. (2017). The effects of age on crash risk associated with driver distraction. *International Journal of Epidemiology, 1*(46), 258–65.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine*, *370*(1), 54–59. https://doi.org/10.1056/NEJMsa1204142

Klauer, C., Ollendick, T. H., Ankem, G., & Dingus, T. A. (2017). *Improving driving safety for teenagers with attention deficit and hyperactivity disorder (ADHD).* National Surface Transportation Safety Center for Excellence.

Simons-Morton, B. G., Guo, F., Klauer, S. G., Ehsani, J. P., & Pradhan, A. K. (2014). Keep your eyes on the road: Young driver crash risk increases according to duration of distraction. *Journal of Adolescent Health*, *54*, S61eS67

Simons-Morton, B. G., Klauer, S. G., Ouimet, M. C., Guo, F., Albert, P. S., Lee, S. E., & Dingus, T. A. (2015). Naturalistic Teenage Driving Study: Findings and lessons learned. *Journal of Safety Research, 54*, 41.e29–41. e44. https://doi.org/10.1016/j.jsr.2015.06.010

Simons-Morton, B. G., Ehsani, J. P., Gershon, P., Klauer, S. G., & Dingus, T. A. (2017). Teen driving risk and prevention: Naturalistic driving research contributions and challenges. *Journal of Safety Research* (Special Issue).

Investigating drivers' visual scanning patterns under different levels of driving automation across daytime and after dark driving environments

Yee Thung Lee[1*], Rafael C. Gonçalves[1], İbrahim Öztürk[1], Karl Proctor[2], Natasha Merat[1]

[1] Institute for Transport Studies, University of Leeds, Leeds, United Kingdom
[2] Jaguar Land Rover, Coventry, United Kingdom
[*]tsytl@leeds.ac.uk

Abstract: Using a within-participants design, we investigated how visibility of the driving scene (daytime, after dark) and level of automation (manual driving, L2 automation) affected drivers' visual scanning patterns. Fourty four paricipants took part in this driving simulator study, which incorporated an urban environment that included vulnerable road users (VRUs) and a range of roadside elements. We found that the pattern of drivers' horizontal gaze dispersion was similar for manual and L2 driving, suggesting that visual attention towards VRUs and roadside elements was the same, regardless of automation level. Participants had a higher vertical gaze dispersion during manual driving, compared to L2 automation, reflecting increased demands for maintaining the designated speed limit in the former condition. Lighting condition did not significantly affect either horizontal or vertical gaze dispersion. However, a significant interaction was found for vertical gaze dispersion between lighting condition and automation level. Although L2 automation reduced vertical gaze dispersion compared to manual driving for both lighting conditions, the reduction was more pronounced in daytime conditions. This suggests that reduced visibility of the driving environment makes it challenging for drivers to detect VRUs and other traffic elements at a further distance, thereby weakening the advantages provided by L2 automation for attention management, caused by the reduction of visual attention demands associated with the vehicle's operational control.

1. Introduction

1.1 The effect of SAE Level 2 automation on gaze patterns

Commercially available SAE Level 2 (L2) automated vehicles are equipped with a driving assistance system that allows drivers to delegate lateral and longitudinal control of the driving task to the vehicle (SAE, 2021). When L2 is engaged, the reduction of visual attention demands associated with the vehicle's operational control may allow drivers to horizontally scan the driving environment more extensively (Louw & Merat, 2017; Mackenzie & Harris, 2015). Other study have shown that vertical gaze dispersion was also higher in manual driving than L2 automation, as driver constantly shifted their gaze between forward roadway and instrument cluster (Goncalves et al., 2020).

1.2 Driving after dark

Low visibility of the driving scene after the hours of darkness is associated with increased collision rates (Fotios et al., 2021; Wood, 2020). Reduced visibility also impairs peripheral hazard detection (Garay et al., 2004; Garay-Vega et al., 2007). Previous driving simulator studies, using manual driving, have not shown significant differences in visual scanning patterns between daytime and after-dark driving environments (for both horizontal and vertical gaze dispersions; Konstantopoulos et al., 2010). Regarding real world studies, however, a review by Grüner and Ansorge (2017) found reductions in horizontal gaze dispersion after dark, compared to daytime driving.

1.3 The present study

To date, studies on drivers' visual scanning patterns in L2 automation have mainly used motorway scenarios and focused on daytime driving. Whether these findings are applicable to busy urban environments and in after dark conditions remains unclear. It could be argued that when L2 is engaged, the freed-up resources from operational control could allow more extensive scanning of the driving environments, thereby compensating for the increased visual demands associated with after dark driving in a busy urban environment. Therefore, the aim of the current study was to investigate drivers' visual scanning patterns, comparing manual driving and hands-off L2 automation, in an urban environment across daytime and after dark driving environments.

2. Method

2.1 Participants

44 participants (20 female; 24 male, M_{Age} = 40.23; SD_{Age} = 11.19) participated in this study. They were UK-licensed drivers with at least 3 years' driving experience, and an annual mileage ranging from 3000 to 18000 miles (M = 8125; SD = 3313). The study was granted ethical approval (BESS + FREC 2023 – 0792 – 877) from the University of Leeds.

2.2 Apparatus

The high-fidelity motion-based University of Leeds Driving Simulator (UoLDS) was used for this study. A set of head-mounted Tobii 3 Glasses were used to record eye-tracking data at a rate of 50 Hz.

2.3 Experimental and roadway design

A 2 x 2 within-participants design was implemented with level of automation (manual driving, L2 automation) and lighting conditions of the driving environment (daytime, after dark) as independent variables. Participants completed four drives, and the sequence of drives was counterbalanced across participants.

For manual driving, participants were instructed to drive as they normally would, while adhering to the speed limit (30 mph). For L2 automation, the vehicle was equipped with a hands-off driving assistance system. Although the system was responsible for lane centreing and speed control, participants were told that it could not detect or respond to any hazards in its path. Therefore, they were responsible for monitoring the driving environment at all times and taking over manual control, to avoid a collision.

Different VRUs and roadside elements were placed along the road, every 40 metres (see Fig. 1). These were placed equally on the left and right of the drivers' peripheral vision. Oncoming vehicles travelled at a rate of 1 vehicle per kilometre.



***Fig 1**. Example of VRUs and roadside elements in daytime and after dark driving environments.*

2.4 Procedure

After receiving the experimental briefing and providing informed consent, participants completed two practice drives before the main experimental drives. Each drive lasted 9 minutes, and the total duration of the study was ~ 90 minutes, including the debrief session.

2.5 Data preprocessing and analysis

To control for the influence of road curvature on gaze behaviour, only gaze data from straight road segments with peripheral road elements were considered for data analysis. To control for the influence of hazardous events (e.g., jaywalking pedestrians, see Lee et al., 2026) and the transition of control on gaze behaviour, data from the road segments 100 metres before and after these events were excluded for the analysis. Horizontal and vertical gaze dispersion were used to explore drivers' visual scanning patterns in an urban environment during uninterrupted driving. Mixed-effects models (Stroup, 2012) were used to analyse the data, with backward elimination applied to refine the model, reducing the risk of overfitting (James et al., 2021).

3. Results

Participants' horizontal gaze dispersion was similar for manual driving and L2 automation [F (1, 115.39) = 2.380, p =.126]. In contrast, vertical gaze dispersion was significantly reduced (by ~33 %) in L2 automation, relative to the manual driving conditions [F (1, 118.73) = 245.931, p < .001]. No significant main effect of lighting condition was found for either horizontal [F (1, 114.92) = 0.391, p = .533] or vertical gaze dispersion [F (1, 118.24) = 0.339, p =.562]. No interaction effect was found for horizontal gaze dispersion [F (1, 114.92) = 0.145, p = .704]. However, a significant interaction was found between lighting condition and automation level, when considering vertical gaze dispersion [F (1, 118.24) = 4.722, p = .032]. Post-hoc comparisons showed that while L2 automation significantly reduced the vertical gaze variability compared to manual driving, during daytime and after dark driving, the magnitude of this reduction was significantly more pronounced in daytime (by ~ 37 %) than after dark (by ~ 29 %) (see Fig 2).

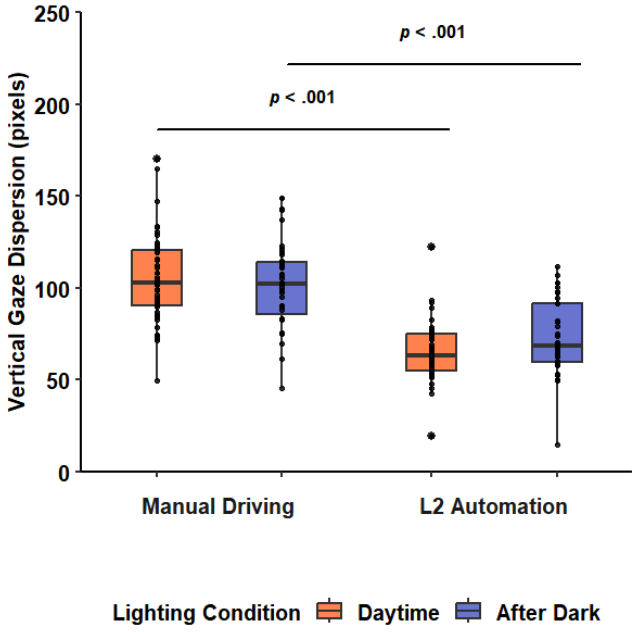


***Fig 2**. Boxplots showing the vertical gaze dispersion across the four experimental conditions.*

4. Discussion and Conclusions

Horizontal gaze dispersion did not differ between manual driving and L2 automation, contradicting previous findings (Louw & Merat, 2017; Mackenzie & Harris, 2015). This could be due to the urban scenarios used in the present study, in contrast to the mainly motorway environment used in previous studies, many of which incorporated car following scenarios, prompting gaze towards the forward roadway. The regular presence of peripheral road elements in the current study attracted drivers' visual attention, causing the high dispersion in horizontal gaze. Consistent with previous findings (Goncalves et al., 2020; Qin et al., 2025), drivers had greater vertical gaze dispersion during manual driving than during L2 automation. This is likely due to the speed-monitoring demands required for manual driving.

Drivers' visual scanning patterns were similar for the daytime and after dark driving environments, consistent with

previous work (Konstantopoulos et al., 2010). This lack of an effect from lighting may be because the simulated setting did not reflect the difficulties that drivers encounter in a real urban nighttime setting (Wood & Chaparro, 2011), for example, the fatigue experienced by the drivers due to circadian rhythm when driving at night (Watling et al., 2022). L2 automation changed drivers' vertical scanning patterns, but this was modulated by the lighting conditions. Specifically, the reduction in vertical gaze dispersion observed in L2 automation was less pronounced in the dark, suggesting that the reduced visibility of the driving environment make it challenging for the drivers to detect the VRUs and other traffic elements from a far distance. Consequently, although L2 automation reduces the necessity for downward gaze shifts towards the instrument cluster for speed-monitoring, drivers still vertically disperse their gaze to detect VRUs and traffic elements, which only become perceivable at a closer distance, necessitating greater vertical scanning during after dark. This ultimately seems to counteract the expected advantageous effects of L2 automation on driver's attention management, caused by the reduction of visual attention demands, associated with the vehicle's operational control.

5. Acknowledgements

This study was one of the experiments funded by the EPSRC HAROLD project (HAzards, Road Lighting and Driving, EP/S003576/1). The vehicle data associated with the hazardous events in this driving simulator study was presented at DDI 2024 with the title "Using a driving simulator study to evaluate drivers' hazard avoidance under different automation levels during day- and night-time conditions". This data is part of a larger study that is currently in preparation for journal publication. The authors would like to thank the University of Leeds Driving Simulator Team for creating the scenarios, especially Dr Albert Solernou, Mr Anthony Horrobin and Mr Jorge Garcia de Pedro. The first author is a PhD student co-funded by Jaguar Land Rover, via an EPSRC iCASE studentship, and this study is part of her PhD project.

References

Fotios, S., Robbins, C., & Uttley, J. (2021). A comparison of approaches for investigating the impact of ambient light on road traffic collisions. *Lighting Research & Technology, 53*(3), 249–261. https://doi.org/10.1177/1477153520924066

Garay, L., Fisher, D. L., & Hancock, K. L. (2004). Effects of driving experience and lighting condition on driving performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*(19), 2290–2294. https://doi.org/10.1177/154193120404801918

Garay-Vega, L., Fisher, D. L., & Pollatsek, A. (2007). Hazard anticipation of novice and experienced drivers: Empirical evaluation on a driving simulator in daytime and nighttime Conditions. *Transportation Research Record*, *2009*(1), 1–7. https://doi.org/10.3141/2009-01

Goncalves, R. C., Louw, T. L., Quaresma, M., Madigan, R., & Merat, N. (2020). The effect of motor control requirements on drivers' eye-gaze pattern during automated driving. *Accident Analysis & Prevention*, *148*, 105788. https://doi.org/10.1016/j.aap.2020.105788

Grüner, M., & Ansorge, U. (2017). Mobile eye tracking during real-world night driving: A selective review of findings and recommendations for future research. *Journal of Eye Movement Research*, *10*(2). https://doi.org/10.16910/jemr.10.2.1

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning : With applications in R*. Springer.

Konstantopoulos, P., Chapman, P., & Crundall, D. (2010). Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. *Accident Analysis and Prevention*, *42*(3), 827–834. https://doi.org/10.1016/j.aap.2009.09.022

Lee, Y. T., Gonçalves, R. C., Ozturk, I., Solernou, A., Proctor, K., & Merat, N. (2026). Hazard perception in manual and hands-off Level 2 driving under daytime and after dark lighting conditions: A driving simulator study. *Cognition, Technology & Work*

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies*, *76*, 35–50. https://doi.org/10.1016/j.trc.2017.01.001

Mackenzie, A. K., & Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Visual Cognition*, *23*(6), 736–757. https://doi.org/10.1080/13506285.2015.1079583

Qin, H., Gonçalves, R. C., Goodridge, C. M., & Merat, N. (2025). Drivers' attention to dash-based human-machine interfaces: The effect of partial automation and cognitive load. *Main Conference Proceedings - 17th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2025*, 198–207. https://doi.org/10.1145/3744333.3747829

SAE. (2021). Summary Of SAE International's Levels Of Driving Automation For On-ROAD Vehicles. Retrieved 29 November 2023, from https://www.sae.org/misc/pdfs/automated_driving.pdf

Stroup, W.W. (2012). Generalized linear mixed models: Modern concepts, methods and applications. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2012, ISBN: 1439815127, 9781439815120

Watling, C. N., Larue, G. S., Wood, J. M., & Black, A. (2022). An on-road examination of daytime and evening driving on rural roads: physiological, subjective, eye gaze, and driving performance outcomes. *Attention, Perception & Psychophysics*, *84*(2), 418–426. https://doi.org/10.3758/s13414-021-02424-9

Wood, J., & Chaparro, A. (2011). Night driving: How low illumination affects driving and the challenges of simulation. *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*, 28-1.

Wood, J. M. (2020). Nighttime driving: Visual, lighting and visibility challenges. *Ophthalmic and Physiological Optics*, *40*(2), 187–201. https://doi.org/10.1111/opo.12659

Anxiety as a Source of Inattention During Highly Automated Driving: Effects on Situation Awareness at Takeover

Angus McKerral [1,2*], Kristen Pammer[1,2]

[1] School of Psychological Sciences, The University of Melbourne, Parkville, VIC, Australia
[2] School of Psychological Sciences, University of Newcastle, Callaghan, NSW, Australia
[*]Angus.McKerral@unimelb.edu.au

Abstract: Anxiety represents an underexamined internal source of attentional disruption in the context of highly automated driving. Whereas non-driving-related task (NDRT) engagement and fatigue are established drivers of inattention and degraded situation awareness (SA) during automation, individual affective states – particularly anxiety – have received comparatively little empirical attention. The present study investigated the relationship between operator anxiety (both state and driving-specific) and SA construction during critical takeover scenarios in a Level 3 simulated automated vehicle, across an extended automated drive. Sixty-seven participants were randomly allocated to a passive vehicle-monitoring condition or a self-selected NDRT engagement condition. SA was assessed via a verbal scene description protocol immediately following two takeover requests, separated by approximately 45 minutes of supervised automated driving. Anxiety was assessed using the State-Trait Anxiety Inventory (STAI-Y) and a modified Driving Cognitions Questionnaire (DCQ-AV). Operators in the monitoring condition showed a significant reduction in driving-specific anxiety post-drive, whereas NDRT-engaged operators did not. Change in state anxiety significantly predicted change in SA across the drive, regardless of condition. Schema-level SA showed the largest decline in the NDRT condition. Findings demonstrate the role of state anxiety as an internal source of inattention that disrupts the cognitive integration underlying safe takeover, with implications for the design of adaptive human-machine interfaces.

1. Introduction

Highly automated vehicles (AVs) fundamentally alter the attentional demands placed on the operator by removing direct control of the driving task and replacing it with a supervisory role (Merat et al., 2019). This transition introduces a well-studied inattention problem: operators disengaged from active driving (i.e. being out-of-the-loop) are vulnerable to degradation in situation awareness (SA), particularly at the point of takeover (de Winter et al., 2014). Research into the sources of this inattention has focused primarily on two factors – engagement in non-driving-related tasks (NDRTs) and passive fatigue associated with monotonous monitoring – both of which delay the reorientation of attention to the driving scene and impair the quality of takeover response (Zhang et al., 2019; Jarosch et al., 2019).

However, inattention is not solely a product of task engagement or fatigue. Internal affective states – particularly anxiety – represent a cognitively demanding source of self-directed attention that may disrupt the construction and maintenance of SA. Processing Efficiency Theory (PET; Eysenck & Calvo, 1992) predicts that anxious individuals divert cognitive resources toward worry, reducing the capacity available for task-relevant processing. In manual driving, anxiety has been shown to both impair and, in some cases, enhance performance, depending on task demand characteristics (Taylor et al., 2008). In low-demand contexts such as passive vehicle monitoring, PET predicts that the impairing effects of anxiety on attentional resources are most likely to be expressed – exactly the conditions under which impoverished SA and takeover performance are of greatest concern.

Despite this theoretical basis, the empirical literature on anxiety and automated driving has focused predominantly on technology acceptance and trust (Smyth et al., 2021; Nordhoff & Lehtonen, 2025; Voinescu et al., 2020), rather than on the real-time attentional and cognitive consequences of anxiety during automated driving. Psychophysiological research links more cognitively demanding NDRTs to reduced eyes-on-road time and narrowed gaze dispersion (Du et al., 2020), while a recent meta-analysis confirms that partial automation increases off-road glances and NDRT engagement (Vasta & Biondi, 2025), suggesting that inattention under automation is both common and multifactorial. Yet individual anxiety – a prevalent affective state in the driving population, with Taylor (2018) reporting that 52% of a community sample endorsed some form of driving anxiety – remains a systematically understudied contributor to attentional disruption in automated contexts.

The present study addresses this gap by investigating whether operator anxiety (state and driving-specific) predicts changes in SA across an extended automated drive, and whether monitoring condition (passive vehicle supervision versus self-selected NDRT engagement) moderates the anxiety-SA relationship. Engagement with NDRTs reflects a central real-world use case of highly automated vehicles; their inclusion as an experimental condition allowed examination of whether anxiety functions as an independent source of SA impairment across both task-engaged and supervisory driving contexts. SA was assessed using a verbal protocol structured according to the Perceptual Cycle Model (PCM), allowing differentiation between information collection and higher-order schema integration, reflecting the construction of a coherent mental model of the driving environment. Critically, the direction of any hypothesised anxiety effect was uncertain: some operators may find relinquishing control anxiety-reducing, freeing attentional resources for SA construction; others may find loss of control anxiety-provoking, amplifying attentional competition with

the driving environment. Given that both NDRTs and anxious rumination are most likely to disrupt higher-order schema integration, it was predicted that schema-related SA would show the greatest sensitivity to both anxiety and NDRT condition.

2. Method

2.1 Participants

Sixty-eight participants (76% women; M age = 26.6 years, SD = 8.6) completed the study, recruited from the University of Newcastle community and through public advertising targeting self-identified anxious drivers. One participant was excluded due to a technical failure, yielding a final sample of N = 67. Participants were randomly allocated to either a no-NDRT condition (n = 34), in which they were required to passively monitor the automated vehicle, or an NDRT-permitted condition (n = 33), in which self-selected NDRT engagement was permitted during the automated drive. All participants held a valid driver licence and had normal or corrected-to-normal vision.

2.2 Procedure

Driving scenarios were constructed in a STISIM M300WS simulator using an automation toolbox (Eriksson, de Winter & Stanton, 2018) that enabled Level 3 automated driving (SAE International, 2021). Participants completed a briefing on the vehicle's automated capabilities followed by two practice drives. The primary trial consisted of an automated drive of approximately 50 minutes interrupted by two takeover request (TOR) events – one approximately five minutes into the automated phase, and a second approximately 45 minutes later. TOR events involved either a simulated truck accident with road debris or a partially obscured construction zone. Each TOR was accompanied by a visual-auditory alert and was followed by an SA probe in response to which participants verbally described the driving scene in the seconds preceding the alert.

2.3 Situation awareness and anxiety

SA verbal responses were transcribed and coded according to PCM categories – World Information (perception of the driving environment), Action (own-vehicle and other-vehicle actions), and Schema (integrated understanding and projected future states) – yielding category-specific count scores for each probe point. Anxiety was assessed at multiple time points using: the trait subscale of the State-Trait Anxiety Inventory (STAI-Y2) prior to the testing session; the state subscale (STAI-Y1) pre- and post-drive; and a modified Driving Cognitions Questionnaire adapted for AV-specific anxiety (DCQ-AV), also administered pre- and post-drive. Change scores (the difference between pre- and post-scores)were computed for all SA and anxiety measures.

3. Results

Multiple regression and paired-samples t-tests (with non-parametric alternatives where appropriate) were employed in the analysis of results.

3.1 Anxiety Scores and NDRT engagement

Operators in the no-NDRT (monitoring) condition showed a significant reduction in AV-specific driving anxiety (DCQ-AV) from pre- to post-drive (W = 302, p = .025), whereas operators in the NDRT-permitted condition showed no significant change (p = .556). State anxiety (STAI-Y1) did not significantly change in either condition.

3.2 SA and NDRT engagement

Summed SA count scores declined significantly in the NDRT-permitted condition (t = 2.85, p = .008) but not in the no-NDRT condition (p = .343). When examined by PCM category, schema-related SA showed the only statistically significant within-condition decline across the drive, and this was specific to the NDRT-permitted condition (M_{pre} = 2.45 to M_{post} = 1.70; t = 2.16, p = .038). No significant changes were observed for World Information or Action categories in either condition.

3.3 SA change and Anxiety Scores

A multiple regression model predicting SA change from STAI-Y1 change and experimental condition was significant (R^2 = 0.173, $F(2, 64)$ = 6.67, p = .002). STAI-Y1 change was a significant predictor of SA change (β = −0.387, p = .001), indicating that operators whose state anxiety declined across the drive showed better-maintained SA at the post-drive takeover, regardless of NDRT condition. Experimental condition was not a significant predictor (p = .139).

4. Discussion

These findings position state anxiety as an internal source of inattention that degrades the quality of SA construction during critical takeover events in highly automated driving. The predictive relationship between anxiety change and SA change – independent of whether operators were engaged with an NDRT – is consistent with PET's account of anxiety as a competitor for attentional resources, reducing the cognitive capacity available for integrating environmental information into a coherent driving schema. This mechanism mirrors the external attentional disruption attributed to NDRT engagement, but operates internally and is therefore less amenable to system-level intervention without knowledge of operator state.

The specificity of schema-level decline to the NDRT condition, and the absence of significant changes in information collection categories, suggests that inattention in automated driving contexts is highly consequential at the level of information integration – the cognitive process that enables operators to project environmental states and plan appropriate responses – rather than at the level of perceptual detection alone. This aligns with broader evidence that SA loss under automation reflects failures of higher-order cognitive processing rather than simply reduced exploration of the driving scene (Endsley, 2021). Notably, this pattern reverses the monitoring-condition SA decline observed by McKerral et al. (2023), tentatively suggesting that the higher baseline arousal associated with anxiety may partially buffer against the passive fatigue that otherwise impairs operators in supervisory roles.

The reduction in driving-specific anxiety observed only in operators who actively monitored the vehicle has practical implications for anxiety-informed HMI design. For prospectively anxious drivers encountering automated systems for the first time, requiring active monitoring may serve to calibrate trust and reduce anxiety, with downstream benefits for SA maintenance – though this benefit must be weighed against the passive fatigue effects associated with supervisory monitoring (Jarosch et al., 2019). Importantly, recent work demonstrates that well-designed warning interfaces can reduce out-of-the-loop anxiety and support attentional re-engagement during takeover (Huang et al., 2024), pointing to a pathway for system-level mitigation.

5. Conclusions

These results demonstrate that anxiety functions as an individual-difference factor mediating the relationship between automated driving and operator attentional allocation. Future research should examine whether real-time physiological monitoring of anxiety state can be used to adapt takeover alert timing and modality for individual operators, and whether interventions targeting driving-specific anxiety translate to improved SA and safer takeover performance in clinical and subclinical populations.

6. Acknowledgements

This research was supported by the Commonwealth through an Australian Government Research Training Program Scholarship [DOI: https://doi.org/10.82133/C42F-K220].

References

de Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F, 27*, 196–217. https://doi.org/10.1016/j.trf.2014.06.016

Du, N., Yang, X., & Zhou, F. (2020). Psychophysiological responses to takeover requests in conditionally automated driving. *Accident Analysis & Prevention, 148*, 105804. https://doi.org/10.1016/j.aap.2020.105804

Endsley, M. R. (2021). A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and SPAM. *Human Factors, 63*, 124–150. https://doi.org/10.1177/0018720819875376

Eriksson, A., de Winter, J., & Stanton, N. A. (2018). A toolbox for automated driving on the STISIM driving simulator. MethodsX, 5, 1073-1088. https://doi.org/10.1016/j.mex.2018.08.003

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion, 6*, 409–434. https://doi.org/10.1080/02699939208409696

Huang, A., Derakhshan, S., Madrid-Carvajal, J., Nezami, F., Wächter, M., Pipa, G., & König, P. (2024). Enhancing safety in autonomous vehicles: The impact of auditory and visual warning signals on driver behavior and situational awareness. *Vehicles, 6*. https://doi.org/10.3390/vehicles6030076

Jarosch, O., Bellem, H., & Bengler, K. (2019). Effects of task-induced fatigue in prolonged conditional automated driving. *Human Factors, 61*, 1186–1199. https://doi.org/10.1177/0018720818816226

McKerral, A., Pammer, K., & Gauld, C. (2023). Supervising the self-driving car: Situation awareness and fatigue during highly automated driving. *Accident Analysis & Prevention, 187*, 107068. https://doi.org/10.1016/j.aap.2023.107068

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., … & Keinath, A. (2019). The "out-of-the-loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology and Work, 21*. https://doi.org/10.1007/s10111-018-0525-8

Nordhoff, S., & Lehtonen, E. (2025). Examining the effect of personality on user acceptance of conditionally automated vehicles. *Scientific Reports, 15*. https://doi.org/10.1038/s41598-024-84776-4

SAE International. (2021). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles J3016_202104*. https://www.sae.org/standards/content/j3016_202104/

Smyth, J., Chen, H., Donzella, V., & Woodman, R. (2021). Public acceptance of driver state monitoring for automated vehicles: Applying the UTAUT framework. *Transportation Research Part F*. https://doi.org/10.1016/j.trf.2021.10.003

Taylor, J. E. (2018). The extent and characteristics of driving anxiety. *Transportation Research Part F, 58*, 70–79. https://doi.org/10.1016/j.trf.2018.05.031

Taylor, J. E., Deane, F. P., & Podd, J. (2008). The relationship between driving anxiety and driving skill: A review of human factors and anxiety-performance theories to clarify future research needs. *New Zealand Journal of Psychology, 37*, 28–37.

Vasta, N., & Biondi, F. (2025). Effect of partially automated driving on mental workload, visual behavior and engagement in non-driving-related tasks: A meta-analysis. *Human Factors, 67*, 962–991. https://doi.org/10.1177/00187208251323132

Voinescu, A., Morgan, P., Alford, C., & Caleb-Solly, P. (2020). The utility of psychological measures in evaluating perceived usability of automated vehicle interfaces: A study with older adults. *Transportation Research Part F, 72*, 244–263. https://doi.org/10.1016/j.trf.2020.05.003

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F, 64*, 285–307. https://doi.org/10.1016/j.trf.2019.04.020

Effects of Automated Driving Style and Visual Countermeasures on Motion Sickness During Visual NDRT

Sarah Schwindt-Drews [1*], Bettina Abendroth[1]

[1] Technical University of Darmstadt, Department of Mechanical Engineering, Institute of Automotive Engineering, Research Group Human Factors & Ergonomics (IAD), Otto-Berndt-Straße 2, D-64287 Darmstadt, Germany
[*]sarah_selina.schwindt@tu-darmstadt.de

Abstract: As vehicle automation increases, engagement in NDRT is expected to rise, while motion sickness remains a barrier to user acceptance. Therefore, this paper presents a driving simulator experiment (N = 28) investigating effects of driving style and visual countermeasures on subjectively and psychophysiologically assessed motion sickness. Results show that an aggressive driving style leads to higher reported motion sickness and peripheral LED-based curve cues as well as HUD-level positioning of reading material each reduce perceived motion sickness. Combining both measures provides no advantage over the individual countermeasures.

1. Introduction

With increasing vehicle automation, engagement in non-driving-related tasks (NDRTs) is rising, but motion sickness remains a barrier to user acceptance (Diels & Bos, 2015). Especially visually demanding NDRTs intensify sensory conflicts between visual and vestibular inputs, a primary cause of motion sickness (Bubb, 2015). Another contributor is the automated driving style, whereby stronger longitudinal and lateral acceleration may worsen nausea symptoms (Karjanto et al., 2022; O'Hanlon & McCauley, 1974). Several studies have shown that a restricted forward view also increases motion sickness, whereas peripheral visual references and favorable display positions can alleviate symptoms (Karjanto et al., 2018; Kuiper et al., 2018; Wu et al., 2025; Xie et al., 2023). These approaches are particularly relevant for modern vehicle interiors with large displays, HUDs and ambient lighting. However, previous studies have largely investigated these measures in isolation or during intermittent visual fixation.

Thus, this study examined the effects of automated driving style and visual countermeasures (LED-based curve cues and HUD-level positioning of NDRT) on motion sickness while performing a task requiring sustained visual fixation. It specifically examined the combination of countermeasures and whether their effects are additive.

It was hypothesized that an aggressive driving style would lead to higher levels of motion sickness than a defensive one (H_1). Furthermore, it was expected that LED-based cues (H_2) and HUD-level text positioning (H_3) would reduce motion sickness. It was further hypothesized that the combination of both (LED+HUD) would result in the lowest level of motion sickness (H_4).

2. Method

2.1 Variables

A within-subjects design was employed, with driving style (defensive vs. aggressive) and countermeasures (none, LED, HUD or LED+HUD) as factors.

The dependent variable, motion sickness, was assessed using subjective and objective measures. Subjective motion sickness was measured using the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993). Objective psychophysiological measures included the skin conductance level (SCL), and heart rate variability (HRV). Both SCL and HRV are well-established stress indicators, with motion sickness associated with increased SCL and reduced HRV (LaCount et al., 2009).

2.2 Simulator and Driving Style

The experiment was conducted in a dynamic driving simulator (seat box on hexapod) reproducing longitudinal and lateral vehicle motions. The virtual environment was displayed across three monitors, providing a 160° horizontal field of view. The driving route featured frequent curves, as well as recurring acceleration, braking and steering maneuvers.

To examine the influence of automated driving style on motion sickness, two driving styles were simulated. In addition to testing the driving style hypothesis, the defensive condition served to identify participants who showed severe symptoms already under lower dynamic provocation and were therefore not considered suitable for repeated aggressive driving exposure.

In the defensive condition, longitudinal acceleration was limited to ±3.6 m/s² and lateral acceleration to ±2.4 m/s². In the aggressive condition, longitudinal acceleration reached ±8.9 m/s² and lateral acceleration reached ±9.6 m/s².

2.3 Non-Driving-Related Task

During the whole fully automated drive, participants were instructed to read texts on a tablet. Reading was used as NDRT because sustained visual fixation is particularly effective at inducing motion sickness according to the sensory conflict theory (Bubb, 2015). To minimize visual perception of the driving environment, participants were instructed to hold the tablet on their lap.

2.4 Countermeasures

The first countermeasure involved LED strips at the left and right side of the seat, within the participants' peripheral field of view. The blue LED flashed (0.2 Hz) on the inner curve side to provide anticipatory motion cues.

The second countermeasure involved positioning the tablet displaying the reading material at HUD height. This allowed peripheral perception of the driving scene during reading.

2.5 Procedure

Six 15-minute experimental drives were conducted across two days. Breaks between drives were adjusted to ensure symptoms had subsided before the next drive. After written informed consent was obtained, demographic data were collected.

On the first day, participants completed a baseline drive (defensive, no NDRT), as well as two drives involving defensive and aggressive automated driving styles while performing the NDRT. On the second day, three aggressive automated driving scenarios with NDRT were conducted, each including one or both countermeasures. Except for the baseline measurement, the order of the drives was randomized within each day (see Fig. 1). The SSQ was completed before and after each drive. SCL and HRV were recorded continuously during the drives.

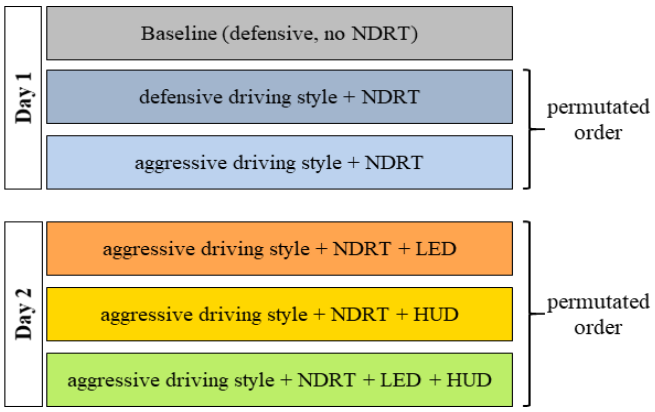


***Fig. 1**. Order of driving conditions.*

2.6 Participants

Data from 28 of 31 participants were analyzed after exclusions due to nonattendance, severe symptoms under defensive driving, or insufficient symptom induction under aggressive driving. The final sample comprised 15 women and 13 men aged between 20 and 63 years (M = 32.96, SD = 13.07).

2.7 Analysis

The SSQ was aggregated into a total score per drive according to the procedure described by Kennedy et al. (1993). Using Ledalab, a SCL value was determined per drive based on electrodermal activity. This was min-max normalized (Lykken & Venables, 1971). The electrocardiogram was analyzed using Kubios to determine a HRV value per drive.

Depending on normality, paired t-tests or Wilcoxon signed-rank tests were used to assess driving style effects (H_1). Effects of the countermeasures (H_2–H_4) were analyzed using repeated-measures ANOVA with Greenhouse–Geisser correction or Friedman tests with Bonferroni-corrected post hoc comparisons.

3. Results

Results are illustrated in Fig. 2. Descriptive statistics are provided in Table 1.

***Table 1** Descriptive statistics (Mean, Standard Deviation).*

	SSQ	normalized SCL [µS]	HRV [bpm]
Baseline (defensive, no NDRT)	M = 292.3 SD = 180.6	M = 0.6 SD = 0.2	M = 5.3 SD = 1.7
defensive driving style + NDRT	M = 432.5 SD = 340.8	M = 0.5 SD = 0.2	M = 4.7 SD = 1.8
aggressive driving style + NDRT	M = 623.0 SD = 453.4	M = 0.5 SD = 0.2	M = 4.8 SD = 1.5
aggressive driving style + NDRT + LED	M = 299.1 SD = 311.3	M = 0.4 SD = 0.2	M = 4.4 SD = 1.7
aggressive driving style + NDRT + HUD	M = 422.0 SD = 361.1	M = 0.6 SD = 0.2	M = 4.6 SD = 1.5
aggressive driving style + NDRT + LED + HUD	M = 350.0 SD = 388.9	M = 0.5 SD = 0.2	M = 4.7 SD = 1.7

3.1 Effect of Driving Style (H_1)

Driving style significantly affected SSQ scores ($t(27) = 6.19$, $p < 0.001$, $d = 1.17$), with higher scores in the aggressive than in the defensive condition. However, no significant differences were observed between driving styles for SCL ($t(27) = -1.05$, $p = .30$, $d = -0.20$) or HRV ($V = 220.5$, $p = .46$, $r = .17$).

3.2 Effect of Countermeasures ($H_{2\text{-}4}$)

A significant effect of the countermeasures was observed for subjectively reported (SSQ) motion sickness (Friedman-$\chi^2(3) = 33.77$, $p < .001$, Kendall's-W = .40). Both the LED ($V = 31$, $p < .001$, $r = -.84$) and HUD ($V = 74$, $p = .017$, $r = -.64$) conditions significantly reduced motion sickness compared to aggressive driving without countermeasures. The combined LED+HUD condition also resulted in a significant reduction ($V = 29$, $p < .001$, $r = -.85$). However, no additional reduction beyond that achieved by the individual measures was observed.

For SCL, a significant overall effect of the countermeasures was found ($F(2.79, 75.32) = 5.11$, $p = .004$, $\eta p^2 = .16$). However, Bonferroni-corrected pairwise comparisons revealed no significant differences between the condition without countermeasures and any of the countermeasure conditions.

Neither the omnibus test (Friedman-$\chi^2(3) = 2.91$, $p = 0.405$, Kendall's-W = 0.03) nor the pairwise comparisons revealed any significant effects of the countermeasures on HRV.

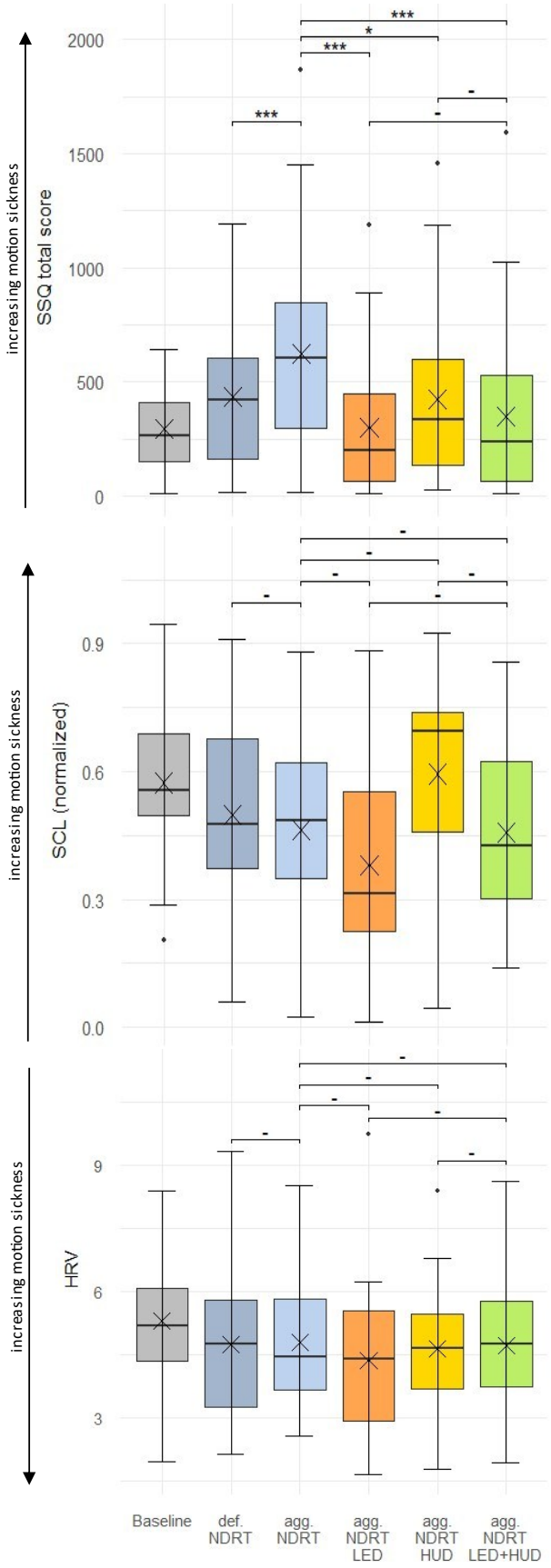


Fig. 2. *Boxplots of the SSQ (top), SCL (middle), and HRV (bottom) across six driving conditions. Horizontal brackets indicate tested pairwise comparisons; asterisks denote significance levels (* $p < .05$, ** $p < .01$, *** $p < .001$), and "–" indicates non-significant differences.*

4. Discussion

Driving style significantly affected subjective motion sickness (H_1), corroborating O'Hanlon and McCauley (1974) and Karjanto et al. (2022). As expected, the aggressive driving style led to significantly higher SSQ scores than defensive driving style. This finding is primarily confirmatory and supports the use of the aggressive driving style as a sufficiently provocative condition for evaluating the visual countermeasures.

Consistent effects of the countermeasures were observed at the subjective level. Both LED-based cues (H_2) and HUD-level text positioning (H_3) significantly reduced perceived motion sickness. These results corroborate those of Karjanto et al. (2018) and Kuiper et al. (2018), demonstrating that both measures are effective during visually demanding NDRTs requiring sustained fixation, even when light cues are implemented more subtly and presented at a greater distance from the occupant.

However, combining both countermeasures (H_4) did not result in any further reduction compared to the individual measures. Instead, the LED+HUD condition showed a tendency towards increased motion sickness, potentially indicating visual overload.

At the psychophysiological level, the effects of driving style and countermeasures were inconsistent. The physiological measures represent non-specific stress indicators and may not detect motion sickness alone.

Given the simulator setting, limited sample size and short exposure duration, generalization to real-world driving and long-term effects requires caution. Future research should investigate the role of individual differences, longer exposure durations and adaptive countermeasure strategies in automated driving contexts.

5. Conclusion

In summary, an aggressive automated driving style increases the subjective experience of motion sickness, whereas visual countermeasures, such as LED-based curve cues or HUD-level text positioning, can effectively reduce it. However, combining both countermeasures did not further reduce motion sickness, suggesting that adding multiple visual aids may not increase effectiveness. The findings therefore extend previous work by demonstrating the effectiveness of visual countermeasures and the limits of combining them during a visually demanding NDRT engagement.

References

Bubb, H. (2015). Gestaltung der Konditionssicherheit. In H. Bubb, K. Bengler, R. E. Grünen, & M. Vollrath (Eds.), *ATZ/MTZ-Fachbuch. Automobilergonomie* (pp. 471–523). Springer Vieweg. https://doi.org/10.1007/978-3-8348-2297-0_8

Diels, C., & Bos, J. E. (2015). User interface considerations to prevent self-driving carsickness. In G. Burnett (Ed.), *ACM Digital Library, Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 14–19). ACM. https://doi.org/10.1145/2809730.2809754

Karjanto, J., Wils, H., Yusof, N., Terken, J., Delbressine, F., & Rauterberg, M. (2022). Measuring the Perception of Comfort in Acceleration Variation Using Electrocardiogram and Self-Rating Measurement for the Passengers of the Automated Vehicle. *Journal of Engineering Science and Technology*(17), 180–196.

Karjanto, J., Yusof, N., Wang, C., Terken, J., Delbressine, F., & Rauterberg, M. (2018). The effect of peripheral visual feedforward system in enhancing situation awareness and mitigating motion sickness in fully automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 678–692. https://doi.org/10.1016/j.trf.2018.06.046

Kennedy, R., Lane, N., Berbaum, K., & Lilienthal, M. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Kuiper, O. X., Bos, J. E., & Diels, C. (2018). Looking forward: In-vehicle auxiliary display positioning affects carsickness. *Applied Ergonomics*, *68*, 169–175. https://doi.org/10.1016/j.apergo.2017.11.002

LaCount, L., Napadow, V., Kuo, B., Park, K., Kim, J., Brown, E., & Barbieri, R. (2009). Dynamic Cardiovagal Response to Motion Sickness: A Point-Process Heart Rate Variability Study. *Computers in Cardiology*, *36*, 49–52.

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology*, *8*(5), 656–672. https://doi.org/10.1111/j.1469-8986.1971.tb00501.x

O'Hanlon, J. F., & McCauley, M. E. (1974). Motion sickness incidence as a function of the frequency and acceleration of vertical sinusoidal motion. *Aerospace Medicine*, *45*(4), 366–369.

Wu, Y., Li, Q., & Chen, X. (2025). Multimodal In-Vehicle Interaction Design to Support Situation Awareness and Reduce Motion Sickness in Autonomous Driving: A Systematic Review. In Q. Gao & J. Zhou (Eds.), *Lecture Notes in Computer Science: Vol. 15811. Human aspects of IT for the aged population: 11th international conference, ITAP 2025 : Held as part of the 27th HCI International Conference, HCII 2025, Gothenburg, Sweden, June 22-27, 2025 : Proceedings, Part III* (Vol. 15811, pp. 120–138). Springer. https://doi.org/10.1007/978-3-031-92712-6_8

Xie, W., He, D., & Wu, G. (2023). Inducers of motion sickness in vehicles: A systematic review of experimental evidence and meta-analysis. *Transportation Research Part F: Traffic Psychology and Behaviour*, *99*, 167–188. https://doi.org/10.1016/j.trf.2023.10.013

Comparison of Driver In-Vehicle Technology Engagement for Hands-On and Hands-Free Partial Driving Automation

T. Zach Noonan [1], Pnina Gershon[1*], Josh Domeyer[2]

[1] Massachusetts Institute of Technology, Center for Transportation and Logistics, AgeLab, Cambridge, MA, USA
[2] Toyota Collaborative Safety Research Center, Ann Arbor, MI, USA
[*] pgershon@mit.edu

Abstract: This study explores how the design of Level 2 driving automation, specifically hands-on versus hands-free design philosophies, is associated with engagement in tasks (driving and non-driving related) involving in-vehicle technologies while driving. Using naturalistic driving data from 31 drivers over a four-week study, it was found that L2 automation use was associated with a higher prevalence of manual cellphone and center stack interactions compared to manual driving (L0). The hands-on system was associated with a 4.45 times greater prevalence of center screen interaction than the hands-free. While drivers using the hands-on system drove with their hands off the wheel 17% of the time, they had 2.75 times greater prevalence of manual cellphone interaction during these periods compared to when their hands were on the steering wheel. In the hands-free system, driving hands off (84% of the time) was associated with 2.38 times greater center screen interaction than with hands on the wheel. Overall, while the evidence suggests that L2 driving automation affords greater engagement with in-vehicle technologies, differences in system design philosophies may shape these interactions.

1. Introduction

The increasing availability of partial driving automation systems (SAE L2) (SAE International, 2016) has coincided with an increase in the amount of in-vehicle technology available to drivers (Consumer Reports, 2021; DeGuzman et al., 2026). As a result, more drivers operate within software-defined and information-rich vehicle environments that include a growing range of interactive interfaces for systems such as infotainment, navigation, and communication (Gershon, 2026). Together, these trends raise questions about driver attention, distraction, and engagement with technologies when using automation.

While L2 systems provide sustained lateral and longitudinal vehicle control, they often differ in their design philosophy and driver-automation interaction strategies. One key difference across commercially available L2 systems is whether the system allows hands-free operation or designed for drivers to keep their hands on the wheel, along with the driver monitoring system (DMS) used to supervise driver engagement. Two such differing systems are GM's Super Cruise (SC) which explicitly allows for hands-free driving and monitors head position/gaze via cameras (Chevrolet, 2026); and Tesla Autopilot (AP) – a hands-on system that (at the time of data collection) monitors the driver via steering wheel torque (Tesla Inc., 2026).

Prior work has shown that vehicle automation use is associated with fewer glances to the forward roadway (Gaspar & Carney, 2019; Gershon et al., 2021), less steering wheel engagement (Morando et al., 2021; Reagan et al., 2021), and greater secondary task engagement (Gershon et al., 2023; Noble et al., 2021; Pipkorn et al., 2025). However, less is known about how the design and implementation of L2 systems shape drivers' engagement with technologies, including portable electronics (e.g., cellphones) and built-in vehicle interfaces such as center screens, center stack controls, steering wheel controls, and hands-free voice systems.

This study uses a large-scale naturalistic dataset – providing real-world data extracted from full trips over a 4 week period of participants driving – to explore whether differences in L2 automation design are associated with differences in driver engagement with in-vehicle technologies. This work addresses two research questions:

- Are differences in the design/implementation of L2 systems associated with differences in the likelihood of technology use?
- Are there differences in the prevalence of technology use between L2 automated systems that explicitly allow hands-free driving and those that require drivers to keep their hands on the wheel when the system is engaged?

2. Method

Data were drawn from the ongoing MIT-AVT naturalistic data collection effort (Fridman et al., 2019; Gershon et al., 2021). 453 trips were recorded from 31 drivers who were randomly assigned a Tesla Model 3 with AP (16 drivers, 233 trips) or a Cadillac CT6 with SC (15 drivers, 220 trips) for 4 weeks. Data were filtered to include highway driving only determined by GPS, automation state was determined by CAN data, and technology interactions and hands-on-wheel (HOW) behaviors were determined from manual annotation of video.

Prevalence was estimated using generalized linear mixed effects models with a negative binomial link function, a total observation time offset term, and a driver-level random effect. A separate model was used for each technology interaction type:

- *Cellphone manual interaction* – manipulating, holding, reaching for, or talking on a hand-held cellphone
- *Center stack interaction* – glancing at the center stack/screen for longer than 1 s, or cumulative 1 s over a 3 s window, or touching the center stack/screen
- *Steering wheel control* – manipulating the controls located on the steering wheel

- *Voice interaction* – using an in-vehicle voice command or talking on a cellphone hands-free

3. Results

Overall, L2 automation use was associated with 2.01 times ($CI_{95\%}$[1.11 – 3.64], p<0.05) higher prevalence of cellphone manual interaction and 1.26 times ($CI_{95\%}$[1.01 – 1.57], p<0.05) higher prevalence of center stack interaction compared to manual driving (L0) (Figure 1). There was no observed difference in steering wheel controls or voice interactions between L2 and L0.

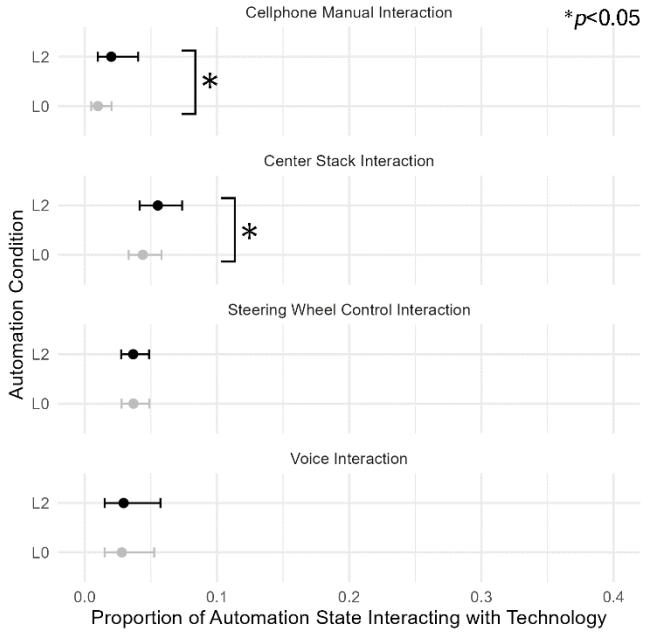


Figure 1. Estimated mean (and 95% CI) proportion of highway driving time in L2 (black) and L0 (gray) in which drivers are interacting with different technologies.

Comparing technology use across L2 systems shows that hands-on L2 system use was associated with a 4.45 times ($CI_{95\%}$[3.30 – 6.01], p<0.001) higher proportion of interaction with the center screen compared to drivers using the hands-free L2 system (Figure 2). There were no significant differences in the prevalence of the other technologies across the two L2 systems.

When using the hands-free L2 system, drivers had their hands off the steering wheel 84% ($CI_{95\%}$ [77.2 – 89.3%]) of the time, while drivers using the hands-on L2 system had their hands off the steering wheel 17% ($CI_{95\%}$ [0 – 43.6%]) of the time. Figure 3 shows the proportion of time interacting with the evaluated technologies, given that drivers either had their hands off or on the steering wheel (at least one hand). When drivers of the hands-on L2 system had no hands on the wheel, manual cellphone interactions were 2.75 times more prevalent than when their hand(s) were on the wheel (CI95% [1.08–7.05], p < 0.05). Drivers of the Hands-Free L2 system interacted with the center screen 2.38 times more frequently when driving hands off the wheel than with hand(s) on the wheel ($CI_{95\%}$[1.53 – 3.72], p<0.001). For both L2 systems, steering wheel control interactions were more prevalent when drivers had hand(s) on the wheel (11.25 $CI_{95\%}$[7.30 – 17.34], p<0.001, and 21.70 $CI_{95\%}$[15.55 – 30.27], p<0.001, respectively).

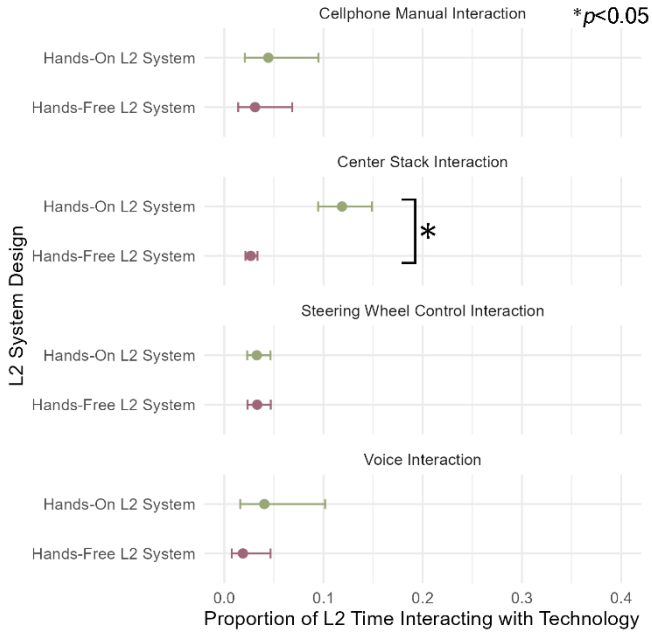


Figure 2. Estimated mean (and 95% CI) proportion of highway driving time in L2 in which drivers are interacting with technology between two different L2 systems.

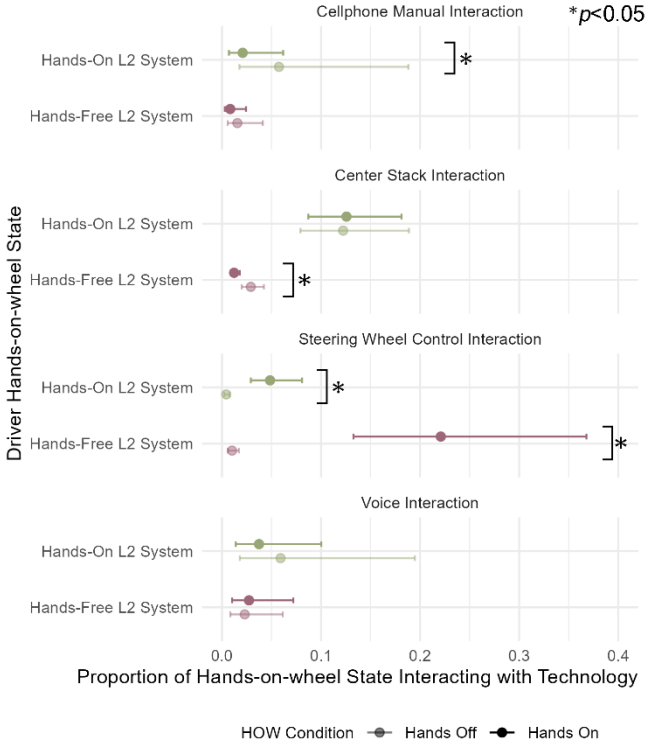


Figure 3. Estimated mean (and 95% CI) proportion of highway driving time in L2 in which drivers are interacting with the different technologies, given that the driver has their hands on the wheel (solid) or hands off the wheel (opaque).

4. Discussion

This study found distinct technology use behaviors when using partial driving automation, including greater prevalence of manual cellphone use and center stack interaction consistent with previous findings (Gershon et al., 2023; Noble et al., 2021). Overall, drivers using the hands-on L2

system had more interaction with the center screen compared to the hand-free L2 system, likely related to both L2 automation and in-cabin physical design. The Tesla Model 3 vehicle has a large center screen and no instrument cluster. As such, drivers are more likely to interact more frequently with the center stack as it contains both driving and non-driving-related information.

Understandably, drivers with the hands-free L2 automation system drove mostly with hands off the wheel, whereas drivers of the hands-on L2 system drove with hands-off the wheel less frequently. However, there were differences in technology use during these hands-off periods. In the hands-free system, drivers interacted more with the center stack and less with steering wheel controls. This may reflect differences in access to similar functions (e.g., infotainment controls) depending on hand placement, or in how they return their hands to the wheel to perform certain tasks (e.g., speed adjustments). In the hands-on system, manual cellphone use was more prevalent when drivers had no hands on the wheel, which, coupled with the similar overall prevalence between both vehicles, may suggest that drivers may rely on the L2 system and remove their hands from the wheel to interact with their cellphones. In contrast, in the hands-free system no difference between HOW behavior, which could be due to the DMS, which is active regardless of the driver's hand position.

5. Conclusions

The goal of this study was to assess differences in partial automation design and drivers' interaction with technologies while driving. Beyond the contribution of physical layout and information displays to the amount of center screen interaction, the key differences between vehicles were observed in how drivers use the system in relation to hands-on-wheel behavior and their engagement with cellphones, center screens, and steering wheel controls. It may be that, while L2 systems generally afford opportunities to engage with technology, the way interactions occur is shaped by the design and implementation of specific systems, including hands-on requirements and monitoring strategies. These findings highlight the importance of understanding driver engagement and distraction when using partial automation systems.

6. Acknowledgments

This research was supported by the Toyota Collaborative Safety Research Center. Data for this study was drawn from work supported by the Advanced Vehicle Technologies (AVT) Consortium at MIT (http://agelab.mit.edu/avt). The views and conclusions expressed are those of the authors and have not been sponsored, approved, or endorsed by supporting organizations.

References

Chevrolet. (2026). *Super Cruise for Select Vehicles*. https://www.chevrolet.com/technology/super-cruise

Consumer Reports. (2021). Understanding the Current State of Vehicle Automation. *Consumer Reports Data Intelligence*. https://data.consumerreports.org/reports/understanding-the-current-state-of-vehicle-automation/

DeGuzman, C. A., Demir, B., Ayas, S., & Donmez, B. (2026). A review of in-vehicle touchscreen safety and usability. *Transportation Research Interdisciplinary Perspectives*, *36*, 101873. https://doi.org/10.1016/j.trip.2026.101873

Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., Terwilliger, J., Patsekin, A., Kindelsberger, J., Ding, L., Seaman, S., Mehler, A., Sipperley, A., Pettinato, A., Seppelt, B. D., Angell, L., Mehler, B., & Reimer, B. (2019). MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction With Automation. *IEEE Access*, *7*, 102021–102038. https://doi.org/10.1109/ACCESS.2019.2926040

Gaspar, J., & Carney, C. (2019). The Effect of Partial Automation on Driver Attention: A Naturalistic Driving Study. *Human Factors*, *61*(8), 1261–1276. https://doi.org/10.1177/0018720819836310

Gershon, P. (2026). Human-aware automation: The future of vehicle intelligence depends on understanding people. *Supply Chain Managment Review*. https://www.scmr.com/article/human-aware-automation-the-future-of-vehicle-intelligence-depends-on-understanding-people

Gershon, P., Seaman, S., Mehler, B., Reimer, B., & Coughlin, J. (2021). Driver behavior and the use of automation in real-world driving. *Accident Analysis & Prevention*, *158*, 106217. https://doi.org/10.1016/j.aap.2021.106217

Gershon, P., Zuby, D., Reagan, I., Schmidt, S., Lenné, M., Fitch, G., & Hernandez, M. (2023). *How Technology Can Help Reduce Driver Distraction*. National Distracted Driving Coalition. USNDDC.org

Morando, A., Gershon, P., Mehler, B., & Reimer, B. (2021). Visual attention and steering wheel control: From engagement to disengagement of Tesla Autopilot. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 1390–1394. https://doi.org/10.1177/1071181321651118

Noble, A. M., Miles, M., Perez, M. A., Guo, F., & Klauer, S. G. (2021). Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. *Accident Analysis & Prevention*, *151*, 105959. https://doi.org/10.1016/j.aap.2020.105959

Pipkorn, L., Zach Noonan, T., Mehler, B., Reimer, B., & Gershon, P. (2025). Changes in glance and hands-on-wheel behaviors when engaging with non-driving-related tasks while using Tesla Autopilot. *Transportation Research Part F: Traffic Psychology and Behaviour*, *109*, 232–241. https://doi.org/10.1016/j.trf.2024.11.004

Reagan, I. J., Teoh, E. R., Cicchino, J. B., Gershon, P., Reimer, B., Mehler, B., & Seppelt, B. (2021). Disengagement from driving when using automation during a 4-week field trial. *Transportation Research Part F: Traffic Psychology and Behaviour*, *82*, 400–411. https://doi.org/10.1016/j.trf.2021.09.010

SAE International. (2016). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (SAE Standard J3016_201609). https://doi.org/10.4271/J3016_201609

Tesla Inc. (2026). *Model Y Owner's Manual* [Concept]. Tesla. https://www.tesla.com/ownersmanual/modely/en_us/GUID-1E76B638-7B12-4D9A-8767-94B7F1E92A0E.html

Evaluating attentional requirements at scale: A GIS-based proof-of-concept

Katja Kircher [1*], Zahra Hamidi [1], Christer Ahlström [1,2]

[1] Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden
[2] Linköping University, Dept of Biomedical Engineering, Linköping, Sweden
[*] katja.kircher@vti.se

Abstract: Within the Minimum Required Attention (MiRA) framework, the areas a driver must monitor are determined by infrastructure, traffic regulations, and the intended manoeuvre. Defining these attentional requirements is typically tedious and largely manual, limiting the scalability of MiRA-based analyses. To examine the feasibility of automating this process, we applied a GIS-based approach to a city block with rich spatial data and tested three scenarios of increasing regulatory complexity. Although the current implementation still involves simplifications, such as excluding signalised intersections, and manual steps, like the identification of attentional requirements based on GIS-derived attributes, the results show that adding dedicated infrastructure and local traffic regulations increases the absolute number of attentional requirements for car drivers globally, while also reducing predictability of what to look out for depending on the intended manoeuvre. Further work is needed to extend the approach to other road users and to evaluate implications for safety, efficiency, and transport justice. With sufficient data quality and an automated requirement-mapping algorithm, GIS-based MiRA analysis has strong potential for larger-scale assessments.

1. Introduction

Road user inattention can be understood as a failure to gather sufficient information to maintain an accurate mental representation of the traffic situation (Endsley, 1995; Johnson-Laird, 1983). In the Minimum Required Attention framework (MiRA; Kircher & Ahlstrom, 2017; Kircher & Ahlström, 2024), this is defined by whether a road user has sampled information from all areas relevant to complete an intended manoeuvre, given the existing infrastructure, traffic rules, and relevant interaction areas. Each manoeuvre is thus associated with a set of *necessary* attentional requirements. These represent the minimum information that must be sampled, assuming all road users obey the rules. For example, when entering an unregulated intersection controlled by priority-from-the-right, one necessary requirement for a driver approaching from a subordinate leg is to check for traffic arriving from the right. Another necessary requirement is to check for bicyclists approaching from behind before making a turn and crossing their path. MiRA also acknowledges *useful* requirements. These represent cases not strictly mandated by the rules, but that are beneficial from a self-preservation perspective. For instance, motorists often fail to notice cyclists approaching from lanes not used by motor vehicles (Ahlström et al., 2025), why MiRA assigns a useful requirement to cyclists as they approach interaction areas involving motor traffic.

A limitation with the MiRA approach has been the time-consuming manual coding of necessary or useful requirements for each manoeuvre and road-user type per infrastructure layout and regulatory context. This manual approach does not scale well in heterogeneous urban environments where intersections vary both in geometry, infrastructure, and rules.

The present paper addresses this methodological gap by initiating the automation of MiRA-based attentional requirement generation using geographic information systems (GIS), exploiting road geometry, pedestrian and cycling facilities, traffic signs, and intersection topologies from municipal and national datasets. This proof-of-concept study is the first step toward generating city wide attention requirement maps, enabling systematic assessment of attentional load across an entire transport network.

2. Method

The assessment was carried out within a geographically defined area in central Linköping, Sweden (Fig. 1). The underlying road network for motor vehicles (centreline data) originated from the Municipality of Linköping, while information on traffic control devices, crossing types, and regulatory signage was primarily sourced from the Swedish National Road Database (NVDB). The latter was supplemented with OpenStreetMap data and

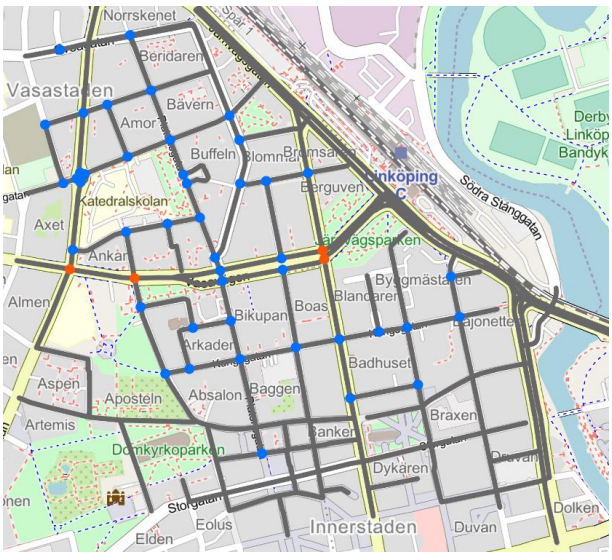


Fig. 1. *The city block used in this proof-of-concept study. Map obtained from OpenStreetMap.*

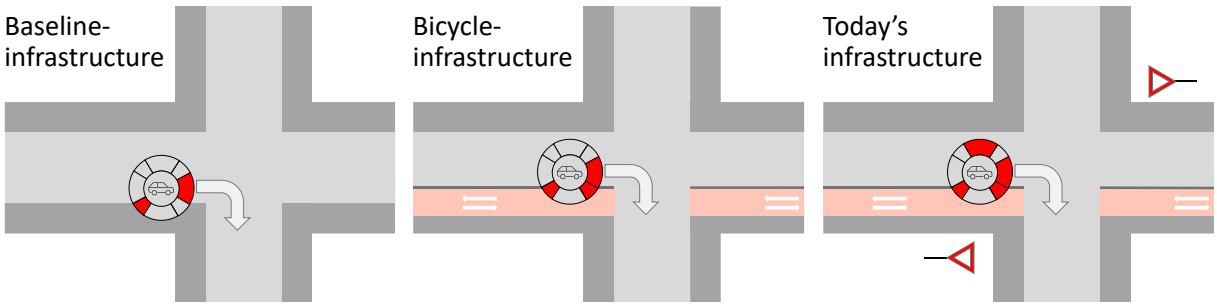


Fig. 2. *Illustration of attentional requirements, depending on the scenario type, when turning right in a four-way intersection. Red fields indicate necessary requirements in that direction.*

Google Street View imagery to validate and correct missing or ambiguous attributes such as turning restrictions, pedestrian crossings, and give-way or priority signs. All spatial analyses were performed in QGIS and ArcGIS Pro, with geometries projected in SWEREF 99 TM to maintain spatial consistency across datasets.

Intersections were categorised according to their geometric configuration, presence of bicycle infrastructure, presence of zebra markings, and traffic regulations or signs. Network topology analysis, supported by spatial joins and buffer-based queries, was used to associate intersection points with connected road segments, and to identify the number of intersection approaches, bicycle paths, and pedestrian crossings. Based on these attributes, a typology of intersections was developed, enabling systematic assignment of MiRA requirements as attributes for structurally similar intersections.

To illustrate the flexibility of the GIS-based approach, three network configurations were constructed (Fig. 2): (i) Baseline (priority-from-the-right) – representing today's road network geometry but without local traffic control devices or dedicated bicycle infrastructure. Only the general rule of priority from the right applied in all unsignalised intersections; (ii) Bicycle-infrastructure scenario – identical to the baseline network except that all existing cycle infrastructure was included. In Sweden, cycle paths are typically bidirectional unless otherwise signed, which affects conflict structure at intersections; (iii) Today's network – representing the actual contemporary conditions, incorporating all zebra crossings, cycle infrastructure, and local regulatory signage (e.g., priority road, give-way, stop), with intersections controlled by traffic lights excluded due to dynamic attentional requirements that vary with signal state. These scenarios serve as illustrative cases for demonstrating how variations in infrastructure and traffic rules can be systematically encoded and subsequently influence the MiRA-derived attentional requirements in the network.

2.1 MiRA requirements

For each intersection leg and manoeuvre, the number and direction of necessary and useful attentional requirements were defined by comparing GIS-derived attributes (priority rules, cycle-path presence, and crossing types) with MiRA's rule logic. Fig. 2 shows an example intersection and how the requirements vary for the three scenarios. In the baseline scenario, for a right turn, in addition to forward, drivers need to check over the right shoulder for cyclists who might go straight on. There are no requirements to look left due to the priority rule, and no requirement to look right since that path is never crossed. Adding a bidirectional cycle path on the right-hand side of the road adds a requirement to also check for oncoming cyclists, while adding a yield rule requires the driver to also check for traffic from the left. Note that the MiRA rule database is still constructed manually, and additional work is needed to algorithmically describe all logical rules.

2.2 Analyses

MiRA requirements were determined for every manoeuvre through each intersection in the study area. This produced a scenario-specific dataset containing, per manoeuvre, the total number of necessary and useful requirements along with the directional distribution of these requirements (over-shoulder right, right, forward-right, forward-left, left, over-shoulder left). Since a requirement to check Forward is always present, the forward requirements were omitted in the analyses. The dataset allow comparison of how changes in infrastructure and rule configurations affect driver attentional demands across the three scenarios.

The analyses presented in this paper are descriptive, aiming to illustrate how the MiRA–GIS framework highlight differences in attentional requirements resulting from changes in infrastructure and regulations. We treat attentional requirements as cumulative (similar to impedance in transport modelling), thereby enabling comparison across scenarios. No inferential statistics are reported as the emphasis is on methodological demonstration rather than differences between the hypothetical scenarios.

3. Results

Fig. 3 shows the number and direction of necessary MiRA requirements across all intersections and manoeuvres in the analysed city block. Note that car drivers have no useful requirements, since MiRA applies useful requirements primarily to pedestrians and cyclists where asymmetric risk exposure justifies additional checks. Across scenarios, the number and directional spread of necessary requirements change systematically. In the baseline scenario, drivers mainly monitor lateral approaches, especially from the right. Introducing bicycle infrastructure adds new potential interaction areas at intersections with cycle paths, increasing necessary checks for turning manoeuvres. In today's network, zebra crossings and local priority rules redistribute attentional demands further. Some checks decrease where pedestrian priority simplifies responsibilities, while others increase where multiple yielding obligations arise. Overall, scenario differences illustrate how infrastructure and regulations shape the attentional environment for drivers.

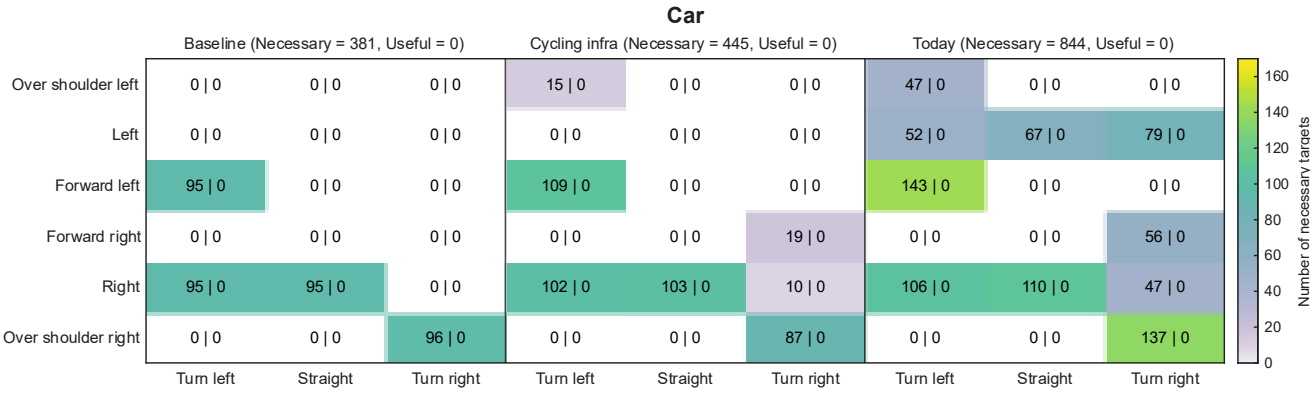


Fig. 3. *Faceted heatmaps of the total number of necessary and useful MiRA requirements in the three road networks (Minimal-control baseline, added cycling infrastructure, today), by attention direction and turn. Each cell is annotated with the total number of necessary | useful requirements*

A single car route through the study area, chosen with route guidance in Google maps, was analysed by summing necessary MiRA requirements at each encountered intersection (baseline requirement n=39, cycling infrastructure n=49, today's network n = 80). The MiRA-GIS approach provides a scalable way to quantify how attentional demands change across environmental scenarios, and can, for example, be used as a parameter in route optimisation.

4. Discussion

In the baseline situation, attentional requirements for car drivers depend solely on the turning manoeuvre. Turning left requires monitoring two distinct target areas (in addition to forward), whereas going straight or turning right requires only one. With added cycling infrastructure and local traffic regulations complexity increases. The number of requirements rises, combinations become less predictable, and required visual scanning can span a wider angle.

Dedicated infrastructure and local rules are often introduced to influence traffic flow or rebalance priority between modes. To gain a complete picture, requirements for other road users would need to be compared directly with those for drivers across scenarios.

Here, only the total number of requirements is presented, but other characteristics, such as the number of simultaneous checks, their spatial distribution, and relative difficulty, should also be examined. These factors may significantly affect attentional demands and the consequences if certain checks are neglected.

The three scenarios (one real and two virtual) demonstrate that mapping attentional requirements across a larger area can reveal systematic patterns and biases linked to risk, efficiency, and fairness between transportation modes. Typical routes could likewise be evaluated and compared within and between traffic modes.

To fully automate MiRA-based requirement mapping, high-quality data on infrastructure layout and local traffic regulations are essential, something which becomes more feasible as digital twins becomes more common. Furthermore, an algorithm capable of using this input to assign attentional requirements for all road user groups and all possible manoeuvres remains to be developed. Future work will focus on this step, alongside the further development of methods for categorising requirements and requirement combinations according to their attentional demands and their potential consequences at a system level.

5. Conclusions

A GIS-based approach to mapping MiRA requirements is technically scalable but depends on the completeness and quality of the underlying data. Several areas of application are conceivable, including traffic planning (e.g., designing for safety or supporting modal shift), research and evaluation of existing and proposed solutions, and the automated-vehicle domain. Future developments include improved assessment of the characteristics of MiRA requirements and their combinations, as well as full automation of the mapping process.

6. Acknowledgments

This work was supported by Länsförsäkringars forskningsfond (Grant Nr. P6/20).

References

Ahlström, C., Kircher, K., Johansson, F., Andersson, A., & Olstam, J. (2025). Driver attention in urban intersections when crossing paths with cyclists. *Accident Analysis & Prevention*, *223*, 108276. https://doi.org/10.1016/j.aap.2025.108276

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human factors*, *37*(1), 32-64. https://doi.org/10.1518/001872095779049543

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Kircher, K., & Ahlstrom, C. (2017). Minimum Required Attention: A Human-Centered Approach to Driver Inattention. *Human factors*, *59*(3), 471-484. https://doi.org/10.1177/0018720816672756

Kircher, K., & Ahlström, C. (2024). A comparison of glance coding approaches for driver attention assessment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *100*, 243-253. https://doi.org/10.1016/j.trf.2023.12.003

Possibility over probability: A worst-case approach to driver distraction and inattention assessment

Abhishek Sarkar [1], Mohammad Ahsan Khodami [1], Juuso-Petteri Partanen [1], Tuomo Kujala [1*]

[1] Cognitive Science, Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland
[*]tuomo.kujala@jyu.fi

Abstract: Conventional approaches to assessing driver distraction and inattention predominantly rely on aggregated task demand and performance measures. Aggregating across conditions may capture average risk but fails to fully represent low-probability, high-consequence events. Understanding how drivers manage risk on the road—and recognizing the catastrophic potential of distracted and inattentive driving—requires that we look beyond the average. We therefore adopt an approach that shifts the focus from what is merely probable to what is possible in worst-case scenarios. This approach offers a principled basis for more robust measurements of distraction and inattention. We operationalize this approach in an empirical study assessing the distraction effects of OEM in-car infotainment systems across ten car models.

1. Introduction

Driving involves operating in situations where even brief moments of distraction or inattention can have severe consequences. The risk posed by driver distraction and inattention is therefore not uniform, but highly context-dependent. Current assessments of driver distraction and inattention are typically based on aggregated task demand and performance measures that emphasize average effects. For example, NHTSA's widely adopted 2-seconds guideline for off-road glances is derived from statistical associations between glance behaviour and safety-critical event risk across large datasets (Klauer et al., 2006; NHTSA, 2013). While useful for establishing baseline safety thresholds, such approaches systematically underweight rare but safety-critical situations.

This is particularly problematic because drivers do not consciously regulate risk but maintain safety margins through largely automatized behavior (Summala, 1988), rendering them unable to adequately compensate when extreme conditions arise. Given the severe consequences of distracted and inattentive driving—including fatalities, injuries, and substantial economic costs—rigorous assessment must prioritize what is possible in worst-case scenarios rather than what is merely probable under typical conditions. This paper therefore advocates a 'possibility over probability' philosophy for assessing driver distraction and inattention, shifting focus from average-based thresholds to worst-case vulnerabilities.

To demonstrate the value of this philosophy in practice, we present an empirical investigation of ten production cars, evaluating the distraction effects of their OEM infotainment systems under worst-case driving scenarios.

2. Method

Kujala and Grahn (2022) proposed a conceptual framework that defines driver inattention as the inability to respond safely to a worst-case scenario—such as a lead car braking abruptly—rather than as a deviation from statistical norms. This ensures that the assessment systematically accounts for improbable but plausible events.

Building on this framework, Kujala and Sarkar (2024) formulated a set of equations to quantify glance- and headway-based measures of driver inattention in dynamic car-following scenarios. The thresholds derived from these equations explicitly account for worst-case scenarios, including sudden braking by the lead vehicle, obstructed peripheral vision, and the inability to swerve to avoid a collision. These thresholds define the boundary between safe driving and collision in a worst-case scenario, capturing risks that would be overlooked by aggregated measures.

These equations provided the foundation for the Drive-In Lab methodology (Kujala et al., 2024; Kujala & Sarkar, 2025), enabling a systematic study of how use of in-car OEM infotainment systems interfere with drivers' ability to maintain safe headway in extreme situations. To further strengthen the worst-case assessment, drivers in the Drive-In Lab use the OEM infotainment systems under study for the first time, with no prior training or familiarization. Kujala and Sarkar (2025) demonstrated this approach in the Drive-In Lab by testing two 2024 production cars (VW ID.7 and Kia EV9), showing how distraction effects can range from negligible to large across different car models and in-car tasks with respect to increase in crash potential under worst-case conditions.

The present study extends the test fleet to ten production cars by adding eight vehicles—seven MY 2024–2025 models (Audi Q6 e-tron, BMW 5 Series, Mercedes-Benz E300e, Peugeot E-3008, Škoda Elroq, Toyota bZ4X, and Volvo EX30) and an older 2016 VW Golf for a comparison point with physical controls for, e.g., AC and seat heating functions. It followed the research design and adopted the operational metrics described by Kujala and Sarkar (2025). Thirty-two participants were recruited per car, with eight drivers in each of four age groups (18–24, 25–39, 40–54, and 55+ years) in line with NHTSA (2013) recommendations. The Drive-In Lab setup and simulation environment were identical to those used in prior testing (Fig. 1).

In all drives, participants' task was to maintain what they think is the minimum safe headway to the lead car by adjusting speed, with automated steering provided but adaptive cruise control disabled. Driving log data were recorded at 10 Hz.



Fig 1. Drive-In Lab's experimental setup from a participant's point of view (car: Audi Q6 e-tron 2025)

Inattention to headway was defined as instances when the driver's distance to the lead vehicle would have been insufficient to avoid a collision if the lead vehicle braked hard suddenly. Distraction was quantified as the extent to which performing an in-car task increased these instances of inattention (%) relative to baseline driving during 3-minute drives.

Statistical analyses were conducted using a linear mixed-effects model, with participant as a random effect and in-car task as a fixed effect, using the baseline drive as the reference. The model estimated the distraction effect of each in-car task on driver inattention with standardized estimates, allowing comparison of distraction across tasks and vehicles.

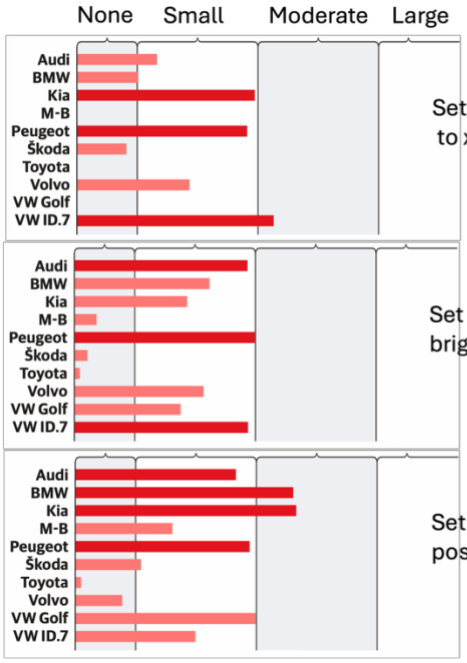


Fig 2. Effect sizes of the least distracting in-car tasks

3. Results

Distraction effect sizes per in-car task and car can be found in Figures 2-4. Figure 2 indicates the least distracting tasks across the cars based on mean effect size (standardized betas) and number of statistically significant effects, Figure 3 shows moderately distracting tasks and Figure 4 the most distracting tasks. Statistically significant effects are illustrated in dark red.

There are notable differences in the distraction effect sizes even in the least distracting tasks between cars. For validation purposes, the car-specific Bonus tasks (Fig. 4) were chosen to represent the most challenging tasks requiring visual search in the menus. The effect sizes of the Bonus tasks support the validity of the novel distraction metric.

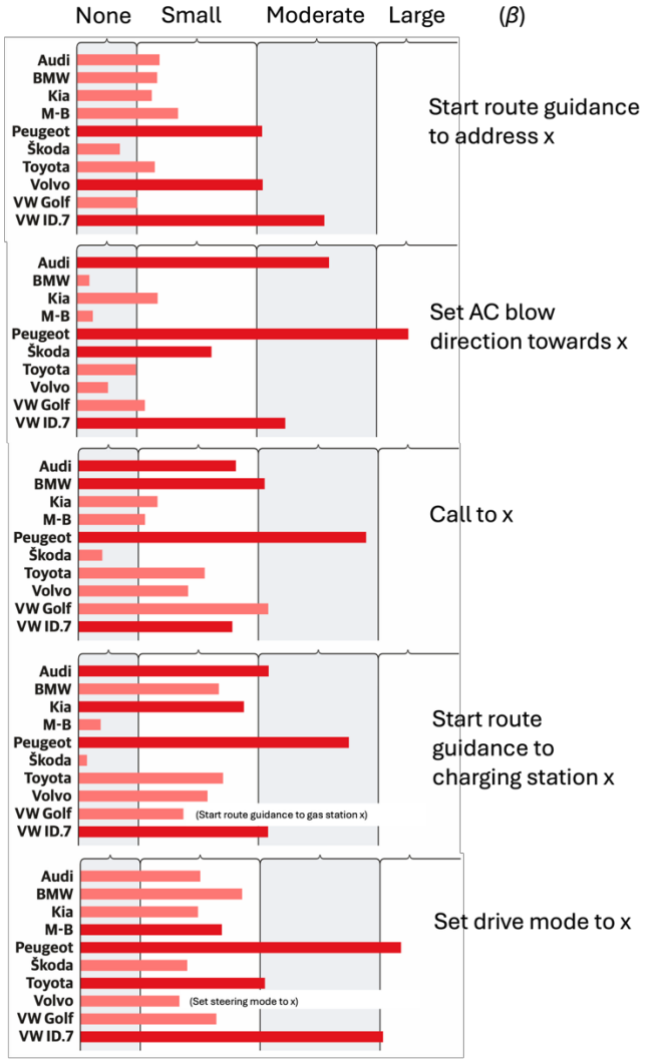


Fig 3. Effect sizes of the moderately distracting in-car tasks (among the studied tasks)

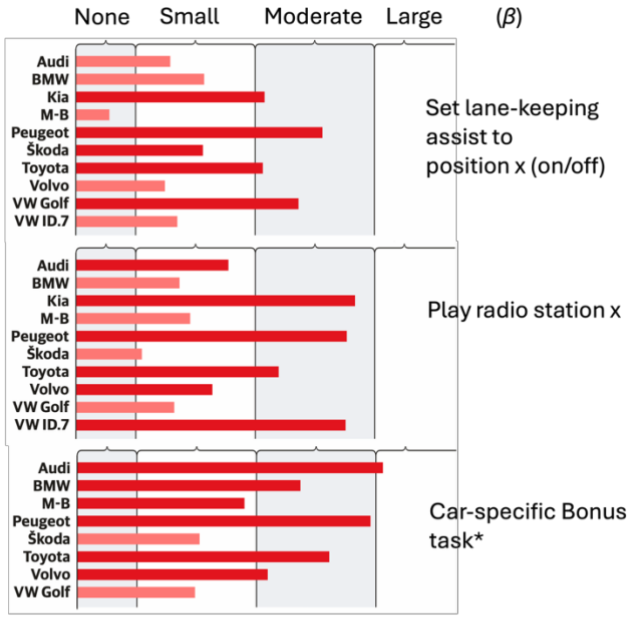
4. Discussion and Conclusion

This paper proposes complementing probabilistic assessments of driver inattention and distraction with evaluations of safety margin maintenance grounded in worst-case possibilities. Literature on risk management (e.g., Taleb et al., 2009) emphasizes that while extreme scenarios cannot be accurately predicted, their consequences must be accounted for to mitigate their impact. Safety-critical systems are therefore often designed around worst-case scenarios, as exemplified in aviation and nuclear power domains. The possibility-based approach is built on the same philosophy: in driving, the costs are fundamentally asymmetric—brief distraction or inattention under typical conditions may go unnoticed, but the same lapse during a worst-case scenario can be catastrophic. Thus, the objective is to minimize the maximum potential loss from distracted and inattentive

driving. While this may yield conservative assessments in some contexts, it reflects the appropriate standard for safety-critical systems where the cost of underestimating risk far exceeds the cost of overestimation. What may appear overly cautious can instead be understood as a minimum safety requirement for preventing crashes caused by distraction and inattention. Prioritizing possibility over probability also enables counterfactual 'what-if' reasoning, defining outcomes based on worst-case conditions to establish causal relationships rather than relying on mere statistical associations (Bärgman et al., 2015; Gerstenberg et al., 2021).

This paper demonstrates an operationalization of the possibility-over-probability philosophy by measuring the worst-case distraction effects of OEM infotainment systems across ten production cars using the Drive-In Lab. The results clearly indicate substantial variation in distraction effects across cars and in-car tasks with effect sizes ranging from negligible to large effects on crash risk potential, underscoring the need for manufacturers to prioritize user interface designs that minimize worst-case vulnerabilities. Effect sizes of the Bonus tasks show construct validity for the distraction measurements. However, the method and measurements must still be further validated in future research. We possess hundreds of hours of video material of drivers' interactions with the in-car systems, and its qualitative content analysis is the next step for revealing the least and most distracting in-car user interface features for a novice user.

Looking ahead, the 'what-if' approach can enhance traffic safety by contributing to the development of robust assessment, monitoring, and warning systems for driver distraction and inattention that ensure protection against extreme scenarios.



*Audi: Set Collision Avoidance Assistant to setting x (on/off), BMW: Set cruise control to setting x, M-B: Set bass level to x (audio), Peugeot: Set mirror folding to setting x, Škoda: Set ESC to setting x (on/off), Toyota: Set speeding alert to setting x, Volvo: Set audio pre-tune to setting x, VW Golf: Set bass level to x (audio)

Fig 4. Effect sizes of the most distracting in-car tasks

5. Acknowledgments

This.research was supported by Research Council of Finland [*Appropriate Uncertainty in Manual and Automated Driving*, grant 343259 and *Drive-In Lab for Testing In-Car User Interfaces*, grant 374023.]

References

Bärgman, J., Lisovskaja, V., Victor, T., Flannagan, C., & Dozza, M. (2015). How does glance behavior influence crash and injury risk? A 'what-if' counterfactual simulation using crashes and near-crashes from SHRP2. *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 152-169.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological review*, *128*(5), 936. https://psycnet.apa.org/doi/10.1037/rev0000281

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

Kujala, T., & Grahn, H. (2022). Key requirements and a method for measuring in-vehicle user interfaces' distraction potential. *The 8th International Conference on Driver Distraction and Inattention*, 60–63. https://ddi2022.org/wp-content/uploads/2022/10/DDI2022-ABSTRACTS.pdf

Kujala, T., Grahn, H., Hautaoja, T., & Sarkar, A. (2024). Drive-In Lab: A method for the measurement and rating of OEM in-car infotainment system's distraction potential. *The 9th International Conference on Driver Distraction and Inattention,* 63-66. https://ddi2024.org/wp-content/uploads/2024/11/DDI2024-ABSTRACTS-.pdf

Kujala, T., & Sarkar, A. (2024). Spare visual capacity and driver inattention in dynamic car following scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour*, *104*, 506-521. https://doi.org/10.1016/j.trf.2024.06.017

Kujala, T., & Sarkar, A. (2025). Evaluating In-Car Tasks' Distraction Effects with Drive-In Lab. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1-24). New York: ACM. https://doi.org/10.1145/3706598.3713590

National Highway Traffic Safety Administration. (2013). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices* (Docket No. NHTSA-2010-0053). https://www.federalregister.gov/d/2013-09883

Summala, H. (1988). Risk control is not risk adjustment: The zero-risk theory of driver behaviour and its implications. *Ergonomics*, *31*(4), 491-506. https://doi.org/10.1080/00140138808966694

Taleb, N. N., Goldstein, D. G., & Spitznagel, M. W. (2009). The six mistakes executives make in risk management. *Harvard Business Review*, 87(10), 78–81.

From glances to outcomes: How counterfactual risk estimates relate to real-world crash odds

Jonas Bärgman [1], Minxiang Zhao [1*], Shweta Joshi [1], Moli Zhou [1]

[1] Chalmers University of Technology, Mechanical Engineering, Hörsalsvägen 7A, Gothenburg, Sweden
[*]minxiang@chalmers.se

Abstract: This study provides a preliminary evaluation of counterfactual, glance-based estimates of model-based crash risk (MCR) and model-based injury risk (MIR) using SHRP2 naturalistic driving data. By comparing MCR with traditional odds-ratio crash estimates, we assess how well prospective, glance-driven predictions align with real-world crash odds. Results indicate that MCR generally reproduces both the ordering and relative magnitude of observed crash risks (R^2 = 0.94 across seven task groups). These findings demonstrate the potential of counterfactual glance-based modeling for prospective safety assessment of DMS-driven inattention countermeasures.

1. Introduction

Driver visual distraction is well established as a major contributor to crash risk (Klauer et al., 2014). However, few methods can prospectively and quantitatively assess the safety impact of distraction before, for example, a human–machine interface (HMI) is deployed in production vehicles. One exception is the counterfactual, simulation-based framework for estimating model-based crash risk (MCR) and model-based injury risk (MIR) from specific glance behaviors, as proposed by Bärgman et al. (2015). This method has been applied to evaluate individual secondary tasks (Bärgman & Victor, 2018), the combined effects of longer glances associated with higher levels of automation and the safety impact of this combination (Bärgman & Victor, 2020), as well as the crash risk associated with drivers' selected time gaps in car-following (Olleja, 2026).

Other quantitative distraction-assessment approaches exist, such as drive-in-lab studies (Kujala & Sarkar, 2025), in which production-vehicle HMIs are evaluated based on drivers' vehicle-control behaviors. However, these methods do not estimate crash or injury risk.

In this work, we (a) estimate MCR and MIR for a range of secondary tasks in the SHRP2 naturalistic driving study (NDS), and (b) compare MCR results with traditional odds-ratio (OR)-based crash-risk estimates and compare MCR/MIR for the same tasks. This comparison serves as an initial validation of the MCR/MIR approach and highlights the importance of considering not only crash risk but also injury risk in the development of driver-monitoring-system (DMS) based inattention countermeasures.

2. Method

2.1 Datasets

2.1.1 The SHRP2 data

We used data from the SHRP2 NDS, in which approximately 3,400 drivers operated instrumented vehicles from 2010–2013, generating about 56 million kilometers of sensor and video data. SHRP2 identified safety-critical events, with crashes classified into four levels; this study uses Levels I–III, and non-critical baseline segments were sampled for comparison. All crash types were included.

Two datasets were analyzed: (1) 6-s crash and baseline epochs for the OR calculations, and (2) 20-s baseline epochs with detailed glance annotations for the MCR/MIR estimates. Secondary tasks were manually coded into 62 categories. In the 567 available 20-s epochs, glance location was annotated every 0.1 s.

For selected secondary tasks, we derived off-road glance distributions from the 6-s task-coding epochs; tasks beginning or ending outside this window contribute only glances within it. Examples of distributions are shown in Figure 1. We fitted parametric models (Weibull or lognormal) to each task's off-road glance durations, accounting for glances cut by the window (Rootzén & Zholud, 2016), and added the share of time with eyes on road (%EON).

For this preliminary work we analyzed OR, MCR and MIR for seven secondary task/task groups (Table 1) – the two task groups include *Visual-manual phone tasks* (Table 2) and *Acceptable tasks* (Table 3). The latter is a group of tasks we believe are societally accepted and unlikely to be legally restricted (e.g., adjusting the radio or climate control, drinking, smoking and talking to passengers). Note that three of the single tasks in Table 1 are also part of a group (radio in the *Acceptable tasks*; texting and phone holding in the *Visual-manual phone tasks*); they are analyzed separately as well, to show their risk individually.

2.1.2 The QUADRIS synthetic crash data

The QUADRIS synthetic crash dataset contains 8,000 pre-crash rear-end kinematic sequences generated and openly

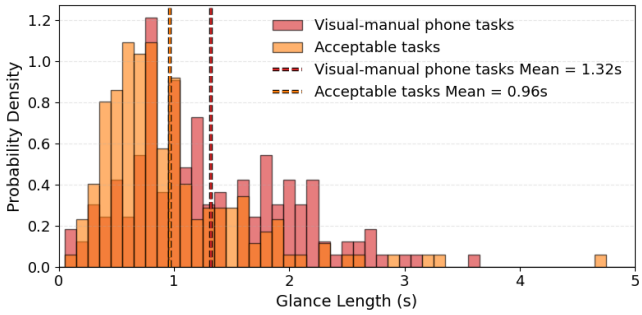


Figure 1: *Distribution of glance lengths for visual-manual phone tasks and acceptable tasks. Dashed lines indicate the mean glance length for each task.*

Table 1 *Secondary tasks/task groups analyzed in this work*

- Visual-manual phone tasks (Table 2)
- Acceptable tasks (Table 3)
- Talking/singing, audience unknown (Talking/singing)
- Phone conversation hand-held (Phone HH)
- Texting; Radio; Phone, holding

Table 2 *Visual-manual phone tasks (SHRP2 annotation categories)*

-Holding	-Texting
-Browsing	-Dialing hand-held
-Locating/reaching /answering	-Dialing hand-held using quick keys
-Other	

Table 3 *Acceptable tasks (SHRP2 annotation categories)*

- Adjusting/monitoring climate control	- Drinking from open container
- Drinking with lid and straw	- Drinking with lid, no straw
- Adjusting/monitoring radio	- Drinking with straw, no lid
- Smoking cigar/cigarette	- Biting nails/cuticles
- Passenger in adjacent seat interaction	- Passenger in rear seat interaction

shared by Wu et al. (2025). The data span the full severity range of real-world rear-end crashes, from minor impacts to fatal events. Each record includes 5 s of kinematic data for both the lead vehicle (LV) and the following vehicle (FV). We used 282 crashes (selected based on absolute and relative speed criteria) as seed crashes in the simulations. To enable counterfactual simulations of FV driver behavior, the driver's evasive maneuver was removed for each seed (see Bärgman et al., 2015).

2.2 Analyses

2.2.1 Odds ratio calculations

To quantify secondary-task effects on crash risk, we used SHRP2 crash and baseline data in propensity- and age-weighted logistic regression models to estimate odds ratios and confidence intervals. For the comparison with MCR, ORs were computed on events in which no task outside the group was coded (task-alone ORs).

2.2.2 Counterfactual crash and injury risk estimates for secondary tasks

Counterfactual simulations were conducted for each secondary task/task group in Table 1 (see VTTI, 2015, for task descriptions). For each seed crash, we combined its pre-crash kinematics with the task-specific glance model, producing a crash-risk estimate for that crash–task pair. Aggregating across seed crashes yielded a crash-risk distribution for each task, from which we computed the expected model-estimated crash risk (MCR).

Injury risk was estimated similarly. For each simulation, we calculated relative impact speed and applied assumptions about crash dynamics together with an injury-risk function (MAIS3+; Wang, 2022). Aggregating across seed crashes produced an injury-risk distribution, and its expected value is reported as the model-estimated injury risk (MIR). Further methodological details for MCR and MIR are provided in Bärgman et al. (2015, 2024).

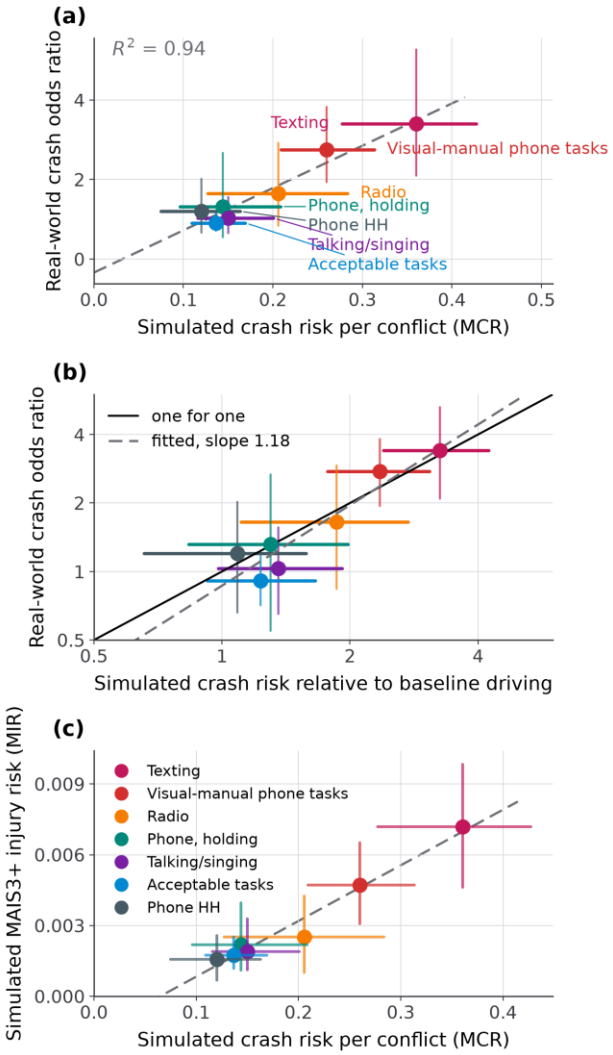


Figure 2: *Simulated risks against real-world crash ORs (task alone) for seven task groups: (a) MCR, with least-squares line; (b) MCR relative to baseline driving (log axes; solid: one for one, dashed: fitted); (c) MCR against MAIS3+ injury risk (MIR). Bars: 95% intervals (horizontal: bootstrap; vertical: OR).*

To estimate the variance of the MCR and MIR values, we applied bootstrapping to both the task windows (refitting the glance models) and the seed crashes. We then conducted two comparisons across secondary tasks: (a) MCR vs. OR, directly and relative to baseline driving, and (b) MCR vs. MIR.

3. Results

3.1 Comparing MCR and OR across secondary tasks

Preliminary results (Figure 2a) show that prospectively estimated MCR relatively closely reproduces both the OR-based ordering of secondary-task risks and their relative magnitudes ($R^2 = 0.94$). For instance, both measures indicate a much larger risk increase for texting and visual-manual phone tasks than for hand-held phone talking/listening, with smaller non-significant differences across other tasks. Further, MCR reflects absolute risk, but it can also be expressed relative to baseline driving, similar to

ORs (Figure 2b). The relative MCRs then match the ORs almost one for one (log-scale slope 1.18).

3.2 Comparing MCR and MIR across secondary tasks

Preliminary results (Figure 2c) indicate that most tasks exhibit substantial variability in both MCR and MIR bootstrap risk estimates. Further, except for texting and the *Visual-manual phone task*, most tasks cluster relatively tightly in both crash and injury risk; the high-risk tasks are more separated from the others in MIR than in MCR.

4. Discussion

This preliminary analysis provides an indication of how well the prospectively estimated, counterfactual glance-based MCR and MIR values correspond to retrospective crash odds observed in the SHRP2 NDS. Comparing these model-derived risks with OR-based estimates across secondary tasks helps assess both the validity and the practical usefulness of the MCR/MIR approach.

4.1 MCR vs. OR

Overall, the results indicate that the counterfactual approach relatively closely reflects the SHRP2 crash odds for several secondary-task glance behaviors, which is promising. The substantial variability is largely due to the limited number of glances per task. The agreement required task-alone ORs. Future work would benefit from larger datasets.

4.2 MCR vs. MIR

The MCR–MIR comparison highlights the distinction between crash risk and injury risk. In a Vision Zero context, DMS-based inattention warnings should therefore prioritize glance behaviors that reduce injury risk (MIR) rather than crash risk alone. The combined MCR/MIR analysis makes these differences clear.

4.3 Limitations and future work

This study has several limitations.

In this preliminary analysis we did not consider total secondary task duration, which affects the risk per task engagement.

The parametric glance models extend beyond the longest observed glances, where the data give little support, so the tail of each distribution depends on the chosen family (Weibull or lognormal).

The current glance-extraction method assigns any glance occurring within a task's start–end window to that task. Consequently, when tasks overlap, all glances—including those from other tasks or from just driving—are attributed to the task being evaluated. It is inherently difficult to disentangle task-specific glances for MCR/MIR calculations.

ORs describe statistical associations rather than underlying crash mechanisms. MCR/MIR, in contrast, estimate crash and injury risk based solely on glance behavior and do not account for contextual driver adaptation, which ORs implicitly capture. Incorporating such adaptation may be feasible, and work by Kujala and Sarkar (2025) and Olleja (2026) may offer guidance.

Further, MCR/MIR is currently limited to rear-end crashes. Extending the approach to additional crash types should be considered, although doing so is not straightforward.

5. Conclusions

Our preliminary results show that counterfactual glance-based estimates capture real-world crash-risk patterns and reveal important distinctions between crash and injury risk. These findings support the MCR/MIR framework as a promising tool for prospective assessment of secondary-task safety and for guiding DMS-based countermeasure designs aligned with Vision Zero injury-prevention goals.

6. Acknowledgments

This work was conducted as part of the coursework for the Traffic Safety Epidemiology course at Chalmers University of Technology, but the work is connected to the QUADRARUM project. This research used data from the SHRP2 Naturalistic Driving Study, hosted by the Virginia Tech Transportation Institute under contract to the Transportation Research Board. The data are accessed under license SHRP2-DUL-A-3-19-432. The findings and conclusions are those of the authors and do not necessarily represent the views of VTTI, TRB, or the National Academies.

References

Bärgman, J., Lisovskaja, V., Victor, T., Flannagan, C., & Dozza, M. (2015). *How does glance behavior influence crash and injury risk? A 'what-if' counterfactual simulation using crashes and near-crashes from SHRP2.* Transportation Research Part F: Traffic Psychology and Behaviour, 35, 152–169. https://doi.org/10.1016/j.trf.2015.10.011

Bärgman, J., Svärd, M., Lundell, S., & Hartelius, E. (2024). *Methodological challenges of scenario generation validation: A rear-end crash-causation model for virtual safety assessment.* Transportation Research Part F: Traffic Psychology and Behaviour, 104, 374–410. https://doi.org/10.1016/j.trf.2024.04.007

Bärgman, J., & Victor, T. (2018, October 15–17). *Using counterfactual simulations to evaluate the impact of drivers' glance behaviors on safety: A study of between-driver variability.* In *Proceedings of the 6th International Conference on Driver Distraction and Inattention*, Gothenburg, Sweden.

Bärgman, J., & Victor, T. (2020). *Holistic assessment of driver assistance systems: How can systems be assessed with respect to how they impact glance behaviour and collision avoidance?* IET Intelligent Transport Systems, 14(9), 1058–1067. https://digital-library.theiet.org/content/journals/10.1049/iet-its.2018.5550

Edvardsson Björnberg, K., Hansson, S. O., Belin, M.-Å., & Tingval, C. (Eds.). (2023). *The Vision Zero handbook: Theory, technology and management for a zero-casualty policy.* Springer. https://doi.org/10.1007/978-3-030-76505-7

Hankey, J. M., Perez, M. A., & McClafferty, J. A. (2016). *Description of the SHRP2 naturalistic database and the crash, near-crash, and baseline data sets—Task report.* https://vtechworks.lib.vt.edu/bitstream/handle/10919/70850/SHRP_2_CrashNearCrashBaselineReport_4-25-16.pdf

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving

and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine, 370*(1), 54–59. https://doi.org/10.1056/NEJMsa1204142

Kujala, T., & Sarkar, A. (2025). Evaluating in-car tasks' distraction effects with Drive-In Lab. In *CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–24). ACM. https://doi.org/10.1145/3706598.3713590

Olleja, P. (2026). *Driving behavior and safety targets: A naturalistic perspective* (Doctoral dissertation, Chalmers University of Technology). Chalmers University of Technology. https://doi.org/10.63959/chalmers.dt/5837

Rootzén, H., & Zholud, D. (2016). *Tail estimation for window-censored processes.* Technometrics, 58(1), 95–103. https://doi.org/10.1080/00401706.2014.995834

VTTI. (2015). *Researcher dictionary for Safety Critical Event Video Reduction Data* (Version 4.1). Virginia Tech Transportation Institute. https://vtechworks.lib.vt.edu/server/api/core/bitstreams/356707d0-e639-4df1-839d-d2cb978374f8/content

Wang, J. S. (2022). *MAIS (05/08) injury probability curves as functions of delta-V* (DOT HS 813 219). https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813219

Wu, J., Flannagan, C., Sander, U., & Bärgman, J. (2025). *Model-based generation of representative rear-end crash scenarios across the full severity range using pre-crash data.* IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2025.3573386

Predicting Reaction Time Under L2 Assisted Driving and Cognitive Load: A Sensitivity Analysis of Eye-Tracking Features Using XGBoost and Explainable AI

Hao Qin [1*], Rafael C. Gonçalves [1], Courtney M. Goodridge [1,2], Natasha Merat [1]

[1] Institute for Transport Studies, University of Leeds, Leeds, United Kingdom
[2] School of Psychology, University of Leeds, Leeds, United Kingdom
[*]tshq@leeds.ac.uk

Abstract: There is evidence that non-visual yet cognitively demanding tasks (e.g., hands-free phone conversations) can impair drivers' reactions to traffic hazards in both manual and Level 2 assisted driving (L2 AD). Monitoring cognitive load and achieving a timely intervention is therefore critical. Eye-tracking features have shown promise for estimating cognitive load, but research has focused on classifying levels of load, with less attention to how these features relate to driver performance changes under this load. Using data from a simulator study, we examined the sensitivity of eye-tracking features for predicting driver reaction times to critical hazards (pedestrians and vehicles) in an urban environment under varying cognitive load (No Task, 2-Back) and driving modes (manual driving, L2 AD). We applied XGBoost to model non-linear relationships and SHAP for interpretability. Results showed that combining eye-tracking features from a hazard visibility window (after hazard appearance) and pre-visibility window (before hazard appearance) improved prediction, compared to using either window alone. SHAP analysis indicated that patterns reflecting lower cognitive load (e.g., lower pupil diameter variance, greater fixation dispersion) before hazard visibility and higher cognitive load (e.g., greater pupil dilation, longer fixation duration) after hazard visibility were associated with faster reactions. This reflects the need to distinguish cognitive load from non-driving tasks and load related to driving and hazard perception. Automation level showed a negligible contribution to prediction. These findings provide guidance for monitoring of non-visual driver distraction in increasingly automated vehicles.

1. Introduction

Level 2 assisted driving (L2 AD) is becoming increasingly available in consumer vehicles (SAE International, 2021). Despite a requirement to monitor the system, research shows that drivers are prone to engaging in non-driving-related tasks (NDRTs) when L2 is engaged (Klauer et al., 2023). The risk of engaging in visual-manual NDRTs (e.g., texting) has being widely acknowledged during the past 15-20 years, while research on the effect of non-visual NDRTs shows mixed results (e.g., hands-free phone conversations) in the context of L2 (Goodridge et al., 2026b). There is evidence that non-visual NDRTs may particularly impair driver reactions in cognitively demanding situations, such as during silent system failures in demanding traffic conditions (Hwang et al., 2024; Qin et al., 2026).

In principal, the safety risks associated with non-visual NDRTs may be mitigated through a driver monitoring system capable of detecting high cognitive load and providing timely interventions. Eye-tracking, in particular, demonstrates potential in this regard, given its established sensitivity to cognitive load, for example through measures of gaze dispersion (Victor et al., 2005) and pupil dilation (Goodridge et al., 2026a). Most studies have validated the sensitivity of eye-tracking metrics to different levels of load (e.g., 0-, 1- or 2-back; Goodridge et al., 2024; Pfleging et al., 2016; Victor et al., 2005). A smaller number of studies have investigated the relationship between eye-tracking metrics and changes in driver performance (e.g., reaction to hazards) (Goodridge et al., 2025, 2026a). However, these attempts relied on linear models, precluding the possibility of investigating non-linear trends between driver state and performance.

Machine learning approaches have been applied to model driver reactions using eye-tracking features and capture such non-linear relationships. However, studies have either focused on manual driving (Liang et al., 2007) or used visual-manual NDRTs (Chen et al., 2024; Du et al., 2020). Gonçalves et al. (2025) examined how takeover time in L2 AD is affected by visual occlusion or cognitive load. However, their analysis was limited to lead-vehicle braking events under takeover requests, offering limited insight into more cognitively demanding scenarios, such as silent failures in the presence of diverse hazards associated with complex urban driving.

The present study examined the sensitivity of eye-tracking features to drivers' reaction times across different hazard types (pedestrians and vehicles) during both manual driving and silent failures of L2 AD. XGBoost with SHAP explanations was employed to capture nonlinear relationships and ensure interpretability.

2. Methods

2.1 Participants

We used data from a previous driving simulator experiment (Qin et al., 2025), involving 46 UK drivers. Participant demographics are summarized in Table 1.

Table 1 *Participant demographics*

Variable	Descriptive Statistics
Age (years)	M = 39, range = 23-64
Driving experience (years licensed)	M = 17, range = 2-45
Annual mileage (miles)	M = 7,197, range = 2,000–18,000
Gender	25 male, 21 female

2.2 Apparatus

The experiment was conducted using a static driving simulator (Fig. 1) equipped with a Smart Eye Pro (v10.0) eye-tracking system sampling at 60 Hz.



Fig. 1. *Driving simulator used in the experiment*

2.3 Design

The study used a 2 × 2 within-participant design with automation (L2 AD, manual driving [MD]) and cognitive load (No Task, 2-Back) as independent variables, resulting in four experimental conditions, corresponding to four drives per participant.

During L2 AD, participants used a L2 system that maintained a constant speed and lane position but did not respond to hazards or issue takeover requests.

In the 2-back condition, participants continuously performed an auditory 2-back task in which digits (0–9) were presented every 2.25 s, and participants were required to verbally report the digit presented two trials earlier (Mehler et al., 2012).

2.4 Hazardous Events

Participants encountered six events per drive, presented in a counterbalanced order. The analysis reported here focused on three critical events: two pedestrian crossings (from the left and right) and one oncoming vehicle turning across the ego vehicle's path (see Fig. 2). Reaction time was defined as the interval between hazard materialization (pedestrians stepped onto the road or vehicles initiated the turn) and drivers' first braking/steering input. A total of 401 valid observations were retained for analysis.



Fig. 2. *Pedestrian and vehicle hazards*

2.5 Feature Identification

Eye-tracking features sensitive to cognitive load were identified (Table 2). These included metrics capturing visual tunneling (Victor et al., 2005), fixation location (Liang & Lee, 2014), fixation number and duration (Broadbent et al., 2023), saccades (Wu et al., 2019), pupil dilation (Goodridge et al., 2026a), and blink (Liang & Lee, 2014).

Table 2 *Features Used in Predicting Reaction Time*

Type	Abbreviation	Description
Eye tracking features	fixSDx/fixSDy/fixSDe	Standard deviation of fixation position in the horizontal (x), vertical (y), and Euclidean dimensions (cm)
	fixMx/fixMy	Mean fixation position in x and y dimensions (cm)
	fixSEntropy/fixTEntropy	Fixation stationary and transition entropy
	fixNum	Number of fixations
	fixDurMean	Mean fixation duration (ms)
	sAmpMean	Mean saccade amplitude (degrees)
	sVelMean	Mean saccade velocity (degrees/s)
	sDurMean	Mean saccade duration (ms)
	pupilMean	Mean pupil diameter (mm)
	pupilStd	Standard deviation of pupil diameter (mm)
	blinkNum	Blink number
	blinkDur	Mean blink duration (ms)
Demographic features	age, gender, drivingExperience, mileage	Participants' age (year), gender (male = 0; female = 1), years holding driving license (year), and annual mileage (miles)
Situational features	automation	Driving mode (MD = 0; L2 AD = 1)
	eventType	Hazard type (left pedestrian = 0; right pedestrian = 1; oncoming vehicle = 2)
	visibleSpeed	Vehicle speed at hazard onset (m/s)

Demographic and situational factors were also included to account for individual and contextual variability. The cognitive load manipulation was not included, as we aimed to leverage eye-tracking features to capture cognitive load–related changes in reaction times.

2.6 XGBoost and Explainable AI

We used XGBoost to model reaction time, given its effectiveness in handling tabular data and modelling driver reactions (Chen & Guestrin, 2016; Gonçalves et al., 2025). Model performance was evaluated using average root mean squared error (RMSE) from 5-fold cross validation.

Eye-tracking features were summarized over two temporal windows (Fig. 3): a hazard visibility window spanning from when hazards entered visible distance to their materialization, and a pre-visibility window extending up to 30 s prior to hazard visibility. Models were trained using each window separately and the hazard visibility window combined with pre-visibility windows of varying lengths. The best configuration was selected based on averaged performance across five random seeds. SHAP values were computed to identify important features and interpret model (Lundberg & Lee, 2017).

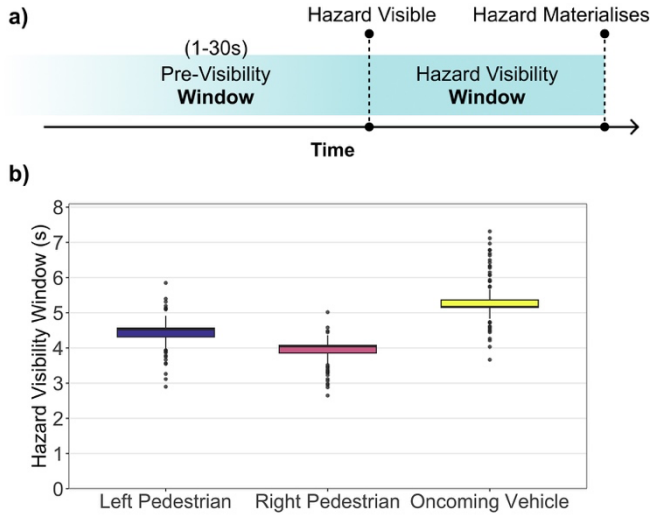


Fig. 3*. Illustration of (a) the time windows and (b) the length of the hazard visibility window.*

3. Results and Discussions

3.1 Model Selection

Optimal cross-seed performance occurred when the hazard visibility window was combined with a 22-s pre-visibility window (Fig. 4). The best model under this configuration achieved RMSE = 0.5624 and R^2 = 0.4972.

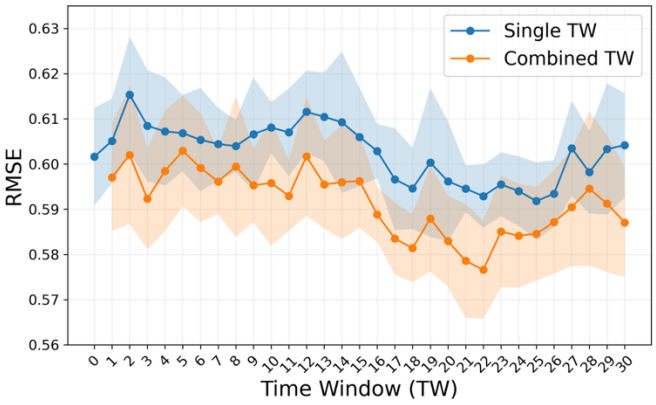


Fig. 4. *Model performance across time windows. Window 0 denotes the hazard visibility window, whereas Windows 1–30 represent the pre-visibility window length (s).*

3.2 SHAP Explaination

The top 15 features ranked by mean absolute SHAP values are presented (Fig. 5). These included situational (e.g., eventType), demographic (e.g., age), and eye-tracking features (e.g., fixDurMean). Automation level ranked among the lowest features and is not shown.

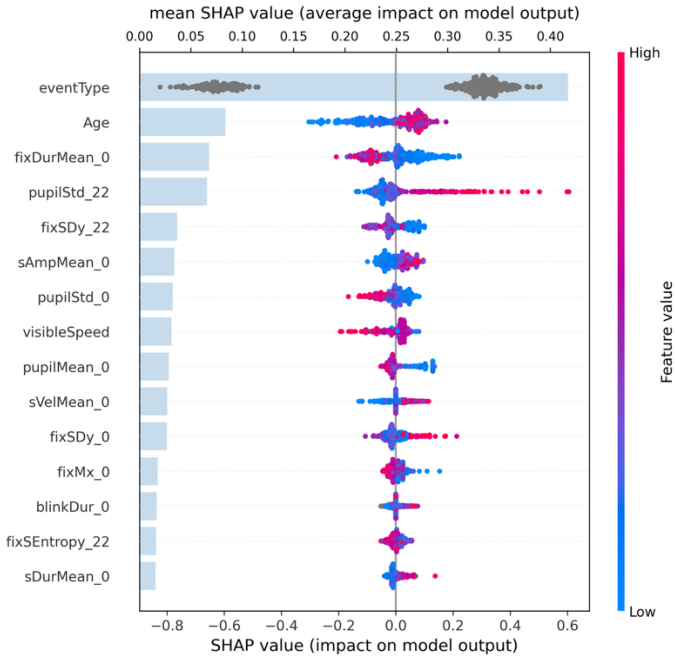


Fig. 5. *SHAP importance and swarm plot (suffix numbers denote the time window).*

We examined SHAP dependence plots (Fig. 6) to further clarify how individual eye-tracking features contributed to the predictions (additional features are shown in the Appendix). For the pre-visibility window, greater pupil diameter variance (Fig. 6b), lower vertical fixation dispersion (Fig. 6c), and lower stationary gaze entropy (Fig. 6k), patterns indicative of higher cognitive load (Goodridge et al., 2024, 2026a), predicted longer reaction time.

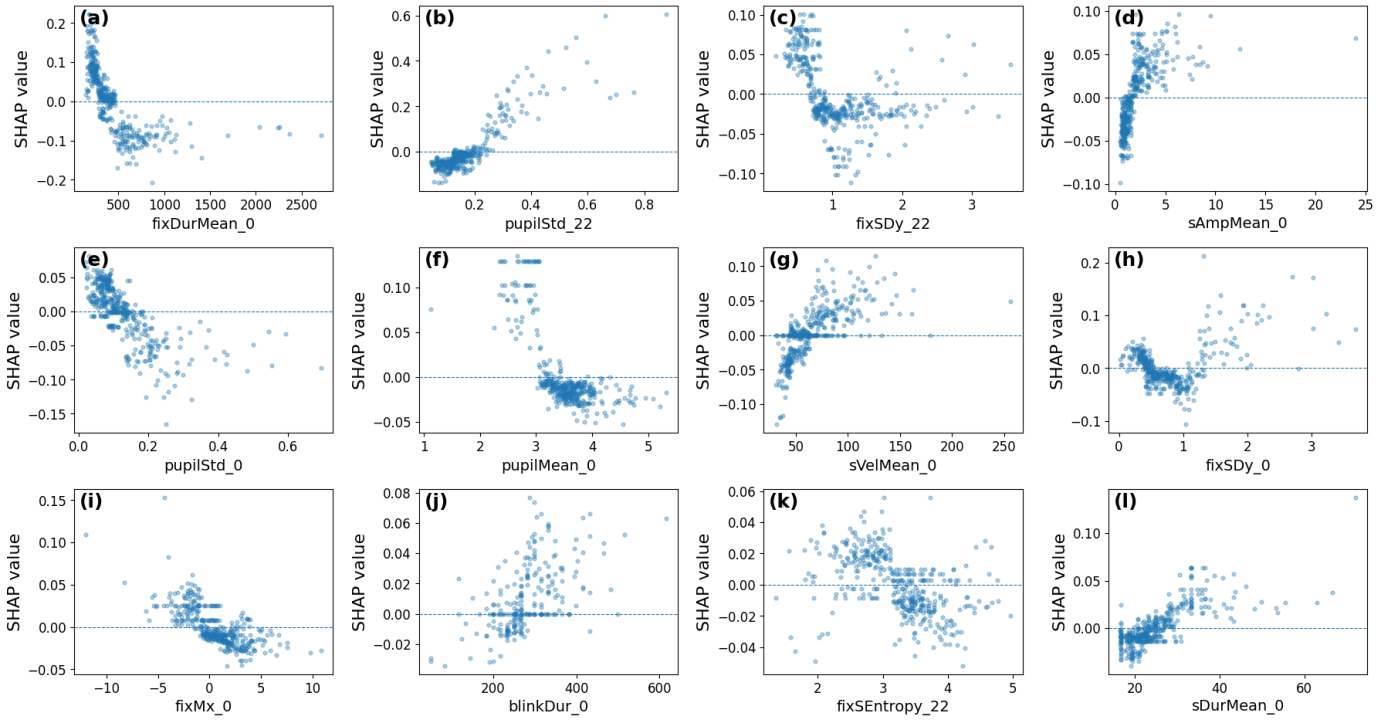


Fig. 5. *SHAP dependence plots (shap values versus feature values).*

During the hazard visibility window, longer mean fixation duration (Fig. 6a), higher mean and variance of pupil diameter (Fig. 6f, 6e), and lower saccade amplitude (Fig. 6d), patterns indicative of increased cognitive load (Broadbent et al., 2023; Radhakrishnan et al., 2022; Wu et al., 2019), were associated with shorter reaction times. These findings suggest that such features may also reflect driving-related effort in hazard perception (Underwood et al., 2011), which may facilitate faster reactions.

The prediction results revealed clear non-linear relationships. For some features (e.g., pupilStd_22), predictions changed sharply once a threshold was exceeded. Others exhibited U-shaped relationships (e.g., fixSDy_22). Further comparisons with linear models are ongoing to determine whether capturing these non-linear patterns improves prediction performance.

4. Conclusions

Our work identified sensitive eye-tracking features related to driver cognitive load that predict reaction times to urban traffic hazards during L2 AD and manual driving. Combining features from both pre- and post-hazard visibility periods proved particularly beneficial. The study could be strengthened by larger samples and the exploration of alternative learners and model linearity, which warrant further work.

5. Acknowledgments

This research was conducted as part of the Hi-Drive project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101006664. The first author was sponsored by a China Scholarship Council–University of Leeds Scholarship.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Broadbent, D. P., D'Innocenzo, G., Ellmers, T. J., Parsler, J., Szameitat, A. J., & Bishop, D. T. (2023). Cognitive load, working memory capacity and driving performance: A preliminary fNIRS and eye tracking study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *92*, 121–132. https://doi.org/10.1016/j.trf.2022.11.013

Chen, H., Zhao, X., Li, H., Gong, J., & Fu, Q. (2024). Predicting driver's takeover time based on individual characteristics, external environment, and situation awareness. *Accident Analysis & Prevention*, *203*, 107601. https://doi.org/10.1016/j.aap.2024.107601

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Du, N., Zhou, F., Pulver, E. M., Tilbury, D. M., Robert, L. P., Pradhan, A. K., & Yang, X. J. (2020). Predicting driver takeover performance in conditionally automated driving. *Accident Analysis & Prevention*, *148*, 105748. https://doi.org/10.1016/j.aap.2020.105748

Gonçalves, R. C., Pardo, J., Elhenawy, M. M. Z., Kuo, J., Azarmi, M., Rezaei, M., Lenné, M. G., Schroeter, R., & Merat, N. (2025). It Matters Who Is Behind The Wheel: Driver Monitoring Feature Analysis Using Explainable AI. *Adjunct Proceedings of the 17th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 245–249. https://doi.org/10.1145/3744335.3758512

Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., Bruneau, A., Kuo, J., Lenné, M. G., Merlhiot, G., Lee, Y. M., & Merat, N. (2026a). *Applying Adaptive Gain Theory to Pupillary Dynamics During Hands-Off L2 Driving Under High Cognitive Load* (Xcw8r_v2). PsyArXiv. https://osf.io/preprints/psyarxiv/xcw8r_v2/

Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., Bruneau, A., Lee, Y. M., & Merat, N. (2026b). The impact of N-back-induced mental workload and time budget on takeover performance. *Accident Analysis & Prevention*, *225*, 108327. https://doi.org/10.1016/j.aap.2025.108327

Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., Lee, Y. M., Madigan, R., & Merat, N. (2024). Gaze entropy metrics for mental workload estimation are heterogenous during hands-off level 2 automation. *Accident Analysis & Prevention*, *202*, 107560. https://doi.org/10.1016/j.aap.2024.107560

Goodridge, C. M., Gonçalves, R. C., Reher, A., Kuo, J., Lenné, M. G., & Merat, N. (2025). Assessing data imbalance correction methods and gaze entropy for collision prediction. *PLOS ONE*, *20*(11), e0336777. https://doi.org/10.1371/journal.pone.0336777

Hwang, J., Choi, W., & Kim, A. (2024). Effects of In-vehicle Auditory Interactions on Takeover Performance in SAE L2 Semi-automated Vehicles. *International Journal of Human-Computer Studies*, 103401. https://doi.org/10.1016/j.ijhcs.2024.103401

Klauer, S. G., Han, S., Guo, F., & Fincannon, T. (2023). Exploring Driver Adaptation to Lower Levels of Automation (L2) Using Existing Naturalistic Driving Data. *27th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration*, (23–0322). https://index.mirasmart.com/27esv/PDFfiles/27ESV-000322.pdf

Liang, Y., & Lee, J. D. (2014). A hybrid Bayesian Network approach to detect driver cognitive distraction. *Transportation Research Part C: Emerging Technologies*, *38*, 146–155. https://doi.org/10.1016/j.trc.2013.10.004

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines. *IEEE Transactions on Intelligent Transportation Systems*, *8*(2), 340–350. https://doi.org/10.1109/TITS.2007.895298

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*. https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Mehler, B., Reimer, B., & Dusek, J. (2011). *MIT AgeLab Delayed Digit Recall Task (n-back)* (SSRN Scholarly Paper No. 6206338). Social Science Research Network. https://doi.org/10.2139/ssrn.6206338

Pfleging, B., Fekety, D. K., Schmidt, A., & Kun, A. L. (2016). A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, 5776–5788. https://doi.org/10.1145/2858036.2858117

Qin, H., Gonçalves, R. C., Goodridge, C. M., & Merat, N. (2026). *When Takeovers Require Cognitive Control: Effects of Partial Automation and Cognitive Load on Driver-Initiated Responses to Traffic Hazards on Urban Roads* (Zr59h_v1). PsyArXiv. https://osf.io/preprints/psyarxiv/zr59h_v1/

Radhakrishnan, V., Merat, N., Louw, T., Gonçalves, R. C., Torrao, G., Lyu, W., Guillen, P. P., & Lenné, M. G. (2022). Physiological indicators of driver workload during car-following scenarios and takeovers in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *87*, 149–163.

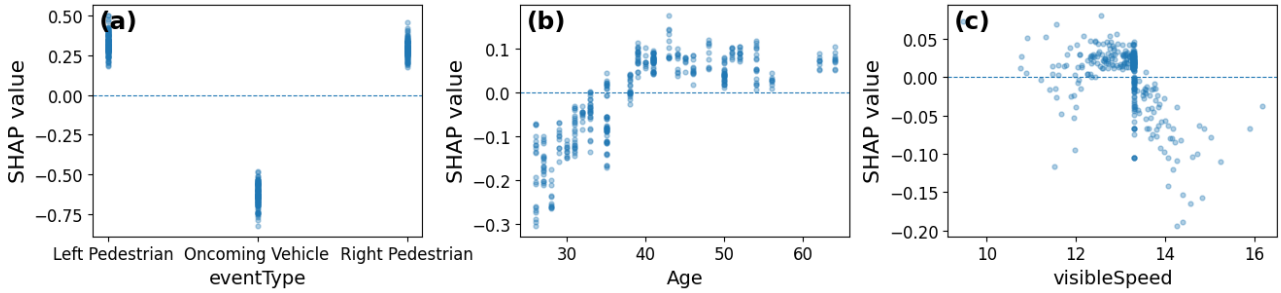
SAE International. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. https://www.sae.org/standards/content/j3016_202104/

Underwood, G., Crundall, D., & Chapman, P. (2011). Driving simulator validation with hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour, Special Issue: Driving Simulation in Traffic Psychology*, *14*(6), 435–446. https://doi.org/10.1016/j.trf.2011.04.008

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 167–190. https://doi.org/10.1016/j.trf.2005.04.014

Wu, Y., Kihara, K., Takeda, Y., Sato, T., Akamatsu, M., & Kitazaki, S. (2019). Assessing the Mental States of Fallback-Ready Drivers in Automated Driving by Electrooculography. *IEEE Intell. Transp. Syst. Conf., ITSC*, 4018–4023. Scopus. https://doi.org/10.1109/ITSC.2019.8917029

Appendix A


(a)
SHAP value
0.50
0.25
0.00
−0.25
−0.50
−0.75
Left Pedestrian
Oncoming Vehicle
Right Pedestrian
eventType
(b)
SHAP value
0.1
0.0
−0.1
−0.2
−0.3
30
40
50
60
Age
(c)
SHAP value
0.05
0.00
−0.05
−0.10
−0.15
−0.20
10
12
14
16
visibleSpeed

10th International Conference
on Driver Distraction and Inattention



DDI2026

GOTHENBURG

October 20-22